

STIC Search Report

Biotech-Chem Library



STIC Database Tracking Number: 125783

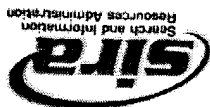
TO: Minh-Tam Davis
Location: REM-3A24&3C18
Art Unit: 1642
Tuesday, June 29, 2004
Case Serial Number: 10/071521

From: Barb O'Bryen
Location: Biotech-Chem Library
Remsen 1A69
Phone: 571-272-2518
barbara.obryen@uspto.gov

Search Notes

RUSH

7/2201



GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 18:01:51 ; Search time 127 Seconds
(without alignments)
404.956 Million cell updates/sec

Title: AAC50492

Perfect score: 852

Sequence: 1 MADEKLPPGWKMRSSG.....GEMSGPVFTDGIHILRTE 163

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_25.*

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_undefined.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	753	88.4	159	13 Q919K6	Q919K6 xenopus lae
2	436	51.2	168	5 Q9N492	Q9N492 caenorhabdi
3	397	46.6	176	3 Q42735	Q42735 emericella
4	369.5	43.4	177	3 Q9C475	Q9C475 candida alb
5	364	42.7	178	3 Q8NUN5	Q8NUN5 cryptococcu
6	351.5	41.3	243	5 P90527	P90527 dictyosteli
7	301	35.3	119	10 Q9SL42	Q9SL42 arabidopsis
8	296.5	34.8	121	10 Q94G00	Q94G00 malus domes
9	283	33.2	118	10 Q9LEK8	Q9LEK8 digitalis l
10	280	32.9	123	10 Q7XTJ8	Q7XTJ8 oryza sativ
11	277	32.5	386	5 Q81R35	Q81R35 drosophila
12	219	25.7	69	10 Q42334	Q42334 arabidopsis
13	218.5	25.6	150	5 Q8SR55	Q8SR55 encephalito
14	185	21.7	630	16 Q82SU8	Q82SU8 nitrosomona
15	181	21.2	246	16 Q899G4	Q899G4 clostridium
16	179.5	21.1	92	1 Q74049	Q74049 cenarchaeum

17	177	20.8	247	16	Q97MB9
18	158	18.5	142	10	Q9FE18
19	157.5	18.5	126	5	Q9NAF9
20	157.5	18.5	130	5	Q9VBU4
21	153.5	18.0	331	16	Q8YUG0
22	153.5	18.0	331	16	Q8FYE0
23	152	17.8	655	16	Q87E47
24	151	17.7	299	16	Q8CXK4
25	150	17.6	93	16	Q7UFM3
26	150	17.6	656	16	Q8PNY3
27	149.5	17.5	248	16	Q8XNR4
28	149.5	17.5	280	16	Q81CB1
29	148.5	17.4	286	16	Q81GY5
30	148.5	17.4	333	16	Q9KDN4
31	147.5	17.3	287	16	Q81U45
32	147	17.3	655	16	Q9PE37
33	146	17.1	656	16	Q8PNI2
34	145	17.0	347	16	Q9JSP0
35	145	17.0	348	16	Q9KJ86
36	143.5	16.8	248	2	Q8GCH6
37	143.5	16.8	283	16	Q81DT1
38	142.5	16.7	434	2	Q7X300
39	140.5	16.5	298	16	Q81QT1
40	138.5	16.3	621	16	Q912T8
41	136.5	16.0	306	16	Q8R760
42	136.5	16.0	325	16	Q8CMR4
43	134.5	15.8	448	16	Q82W17
44	134	15.7	93	16	Q82IA0
45	133.5	15.7	303	16	Q98G68

ALIGNMENTS

RESULT 1
Q919K6
ID Q919K6 PRELIMINARY; PRT; 159 AA.
AC Q919K6
DT 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
DE 01-JUN-2003 (TREMREL. 24, Last annotation update)
DE Prolyl isomerase P1n1.
GN P1N1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20165035; PubMed=10698738;
RA Winkler K.E., Swenson K.I., Kornbluth S., Means A.R.;
RT "Requirement of the prolyl isomerase p1n1 for the replication
RT checkpoint.";
RL Science 287:1644-1647(2000).
DR EMBL; AF239760; AAF43897.1; -.
DR HSSP; Q13526; 1PIN
DR GO; GO:0016853; F.isomerase activity; IEA.
DR InterPro; IPR000297; Rotamase.
DR InterPro; IPR002349; WW.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00639; Rotamase; 1.
DR Pfam; PF00397; WW; 1.
DR PRINTS; PR00403; WWDOMAIN.
DR SMART; SM00456; WW; 1.
DR PROSITE; PS01096; PPIC PPIASE_1; 1.
DR PROSITE; PS01098; PPIC PPIASE_2; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS01159; WW_DOMAIN_2; 1.
KW Isomerase.
SQ SEQUENCE 159 AA; 17671 MW; 4E54F66F16C9DF1C CRC64;
Query Match 88.4%; Score 753; DB 13; Length 159;

```

Db      59 HDGSRNPSSWRSDHITRSKDDANILKSAADLRNPHYEKELKDKANIEGKFKRELAKQFSDC 111
Qy      114 SSARAKGDLGAFGRGQMKQKPFEDASFALRTGEMSGPVFTDSGIHILR 161
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      119 SSARKGGDLGPFERRQMKQKPFEDASFALTEIGEMSDIVDTSSGVHLIYR 166
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
O42735 PRELIMINARY; PRT; 176 AA.
ID O42735
AC O42735;
DT 01-JUN-1998 (TEMBLrel. 06, Created)
DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Peptidyl-prolyl cis/trans isomerase.
GN PINA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RL53;
RX MEDLINE=96151356; PubMed=9482729;
RA Crenshaw D.G., Yang J., Means A.R., Kornbluth S.;
RT "The intact peptidyl-prolyl isomerase, Pin1, interacts with Cdc25 and
RT Plx1.";
RL EMBO J. 17:1315-1327(1998).
DR ENBL; AF035768; AAC49984.1; -.
DR HSP; Q13526; IPIN.
DR GO; GO:0016853; F:isomerase activity; IEA.
DR InterPro; IPR000297; Rotamase.
DR IntexPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00639; Rotamase; 1.
DR Pfam; PF00397; WW; 1.
DR SMART; SM00456; WW; 1.
DR PROSITE; PSS0198; PPIC_PP1ASE_2; 1.
DR PROSITE; PSS01159; WW_DOMAIN_1; 1.
DR PROSITE; PSS00020; WW_DOMAIN_2; 1.
DR KW Isomerase.
SQ SEQUENCE 176 AA; 20038 MW; 1D1E8376239E1309 CRC64;

Query Match 46.6%; Score 397; DB 3; Length 176;
Best Local Similarity 48.0%; Pred.No.2.6e-29;
Matches 82; Conservative 24; Mismatches 51; Indels 14; Gaps 21

Qy      7 LPFGWEKRRSRSSGRVYFYFNHITNASQWRPSPGNS-----SSGGKNGQCEPA-- 53
Db      6 LPAGWEVRHSNKNLPYFNPATRESREWEPPADTDMETLKMVATYHSGAATYHEAPSQE 65
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      54 -RVRCSHLLVKKHSQRRPSSRWREKTRTKEEALALELINGYIQTKSGEEDPESLASQFSD 112
Db      66 GKTRCSHLLVKKHDSRRPSSRWRAEITRTKEEARILRGHQRIMRGIRLGDLMASESD 125
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy      113 CSSAKARGDLGAFSGQMKQKPFEDASFALRTGEMSGPVFTDSGHIILRTE 163
Db      126 CSSARKKGLGPFGRGEMQKFEFEAFALQGVQSDIVESGGLHLIERLQ 176
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
Q9C475 PRELIMINARY; PRT; 177 AA.
ID Q9C475
AC Q9C475;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Peptidyl-prolyl cis/trans isomerase.
OS Candida albicans (yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RA Devasahayam G., Chaturvedi V., Hanes S.D.;
RT "Isolation of the Candida albicans homolog of the ESS1 gene, encoding
RL a prollyl isomerase.", to the EMBL/GenBank/DBJ databases.
RL Submitted (JAN-2000)
DR EMBL; AF224269; AAK00626.1; -
DR HSSP; Q13526; IPIN.
DR GO; GO:0016853; F:isomerase activity; IEA.
DR InterPro; IPR000297; Rotamase.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00639; Rotamase; 1.
DR Pfam; PF00397; WW; 1.
DR SMART; SM00456; WW; 1.
DR PROSITE; PS0198; PPIC_PPIASE_2; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS00020; WW_DOMAIN_2; 1.
DR KX Isomerase.
SQ SEQUENCE 177 AA; 19855 MW; 6AE035D939E778E CRC64;

Query Match 43.4%; Score 369.5; DB 3; Length 177;
Best Local Similarity 43.2%; Pred. No. 9.8e-27;
Matches 73; Conservative 34; Mismatches 49; Indels 13; Gaps 2;

QY 7 LPFGWEKMSRSGRVYFNHITNASOWERPSGN-----SSSGKNGQGEPPAR 54
DB 8 LPFNWTRVRSRHNKEYFLNQSTNESSWDPPYGTDEVLNAYIAKFNKNGYKPLVNEGQ 67
QY 55 VRCSHLLVKSQRRPSSNRO-EKTRTKEALELINGYQIKKSGEDFESLASQFSDC 113
DB 68 VRYSHLLIKNNQKPKSWSPGIGSTRDESQILKXHLERILSGVKULSANTESDC 127
QY 114 SSAAKAGDGLGAFSGQKQKPEDFASALRTGEMSGPVFTDSGHIILRT 162
DB 128 SSHDRGDLGFFSKGQMPPEEAFNHLVHGEVSNIIETNSGVHILQRT 176

RESULT 5
QBUN5 PRELIMINARY; PRT; 178 AA.
AC QBUN5;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Prollyl isomerase Essl.
OS Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX NCBI_TaxID=40410;
RN [1]
RP SEQUENCE FROM N.A.
RA Ren P., Chaturvedi V., Hanes S.D.;
RT "The Ess1 prollyl isomerase is required for virulence in Cryptococcus
RT neoformans.", to the EMBL/GenBank/DBJ databases.
RL Submitted (JUL-2002)
DR EMBL; AF533511; AAN03477.1; -
DR GO; GO:0016853; F:isomerase activity; IEA.
DR InterPro; IPR000297; Rotamase.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00639; Rotamase; 1.
DR Pfam; PF00397; WW; 1.
DR SMART; SM00456; WW; 1.
DR PROSITE; PS0198; PPIC_PPIASE_2; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS00020; WW_DOMAIN_2; 1.
DR KX Isomerase.
SQ SEQUENCE 178 AA; 19529 MW; 6AE3F317E014784B CRC64;

Query Match 42.7%; Score 364; DB 3; Length 178;
Best Local Similarity 45.5%; Pred. No. 3.2e-26;
Matches 81; Conservative 22; Mismatches 45; Indels 30; Gaps 4;

QY 10 GWEKMSRSGRVYFNHITNASOWERPSGNSS-----SGKNGQ 49
DB 5 GWEIRFSNRQIPFYNSERSISTWEPPSELSAEQIQQLFGAAKYNVQLAPAGGKQ 64

```

```

QY 50 GEPARVRCSHLLVKSQRRPSSNROEKTRTKEALELIN---GYIQKIKSGE--EDFE 104
DB 65 -----VRASHILAKHAGSRPASPARNVTRITTSDEQAIIHQHIAVLQSLPPADLPKEFA 119
QY 105 SLASQFSDCSSAKARGDLGAFSGQKQKPEDFASALRTGEMSGPVFTDSGHIILRT 162
DB 120 KIASTEDSCSSARKGGDLGWFGRGQKQKPEDATFNTVPQQLSGIVKTDSGIHILRT 177

RESULT 6
P90527 PRELIMINARY; PRT; 243 AA.
AC P90527;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Pina (Fragment).
GN PINA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelid; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RA Loomis W.F., Iranfar N.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U78757; AAB36960.1; -
DR HSSP; Q13526; IPIN.
DR InterPro; IPR000253; FHA.
DR InterPro; IPR000297; Rotamase.
DR InterPro; IPR008984; SMAD_FHA.
DR Pfam; PF00498; FHA; 1.
DR Pfam; PF00639; Rotamase; 1.
DR SMART; SM00240; FHA; 1.
DR PROSITE; PS00006; FHA_DOMAIN; 1.
DR PROSITE; PS0198; PPI_C_PPIASE_2; 1.
DR NON TER 1
SQ SEQUENCE 243 AA; 26794 MW; 1596001BB784500A CRC64;

Query Match 41.3%; Score 351.5; DB 5; Length 243;
Best Local Similarity 52.8%; Pred. No. 7e-25;
Matches 76; Conservative 15; Mismatches 48; Indels 5; Gaps 2;

QY 18 SSGRVYFNHITNASOWERPSGNSSCGKNGQGEPPARVCSHLLVKSQRRPSSWRQEK 77
DB 103 SSSKXHFILKGTNTN---PS-SSSSSSSSSEPKVTTCRHLVXHQGRNPSSWRESK 157
QY 78 ITRTKEALELINGYQIKKSGEDFESLASQFSDCSSAKARGDLGAFSGQKQKPEDA 137
DB 158 ITRTERATAKINEYRATIIISATFEDLHKNSDCSSAKRGYLDFFKRGQMRPFDC 217
QY 138 SPALRTGEMSGPVFTDSGHIILR 161
DB 218 AFSLVKGEVSGIVDTDSGVHILR 241

RESULT 7
Q9SL42 PRELIMINARY; PRT; 119 AA.
ID Q9SL42;
AC Q9SL42;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative peptidyl-prolyl cis-trans isomerase
DE (AT2G18040/T27K22.9).
DE AT2G18040 OR T27K22.9/AT2G18040.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]

```


OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
 OC Lamiales; Plantaginaceae; Digitalidae; Digitalis.
 OX NCBI_TaxID=49450;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EHRH;
 RC MEDLINE=21216637; PubMed=11118437;
 RA Metzner M., Stollner G., Ruecknagel P., Lu K.P., Fischer G.,
 RA Luckner M., Kuelbertz G.;
 RT "Functional replacement of the essential *esel* in yeast by the plant
 RT parulin *Dlpar13*.";
 RL J. Biol. Chem. 276:13524-13529(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EHRH;
 RC METZNER M.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ133755; CAB94994.1; -;
 DR HSP; Q13526; lPIN.
 DR GO; GO:0016853; F-isomerase activity; IEA.
 DR InterPro; IPR000297; Rotamase.
 DR Pfam; PF00639; Rotamase; 1.
 DR PROSITE; PS01096; PPIC_PP1ASE_1; 1.
 DR PROSITE; PS01098; PPIC_PP1ASE_2; 1.
 KW Isomerase.
 SQ SEQUENCE 118 AA; 12834 MW; 131B74FB4AC3F229 CRC64;
 Query Match 33.2%; Score 283; DB 10; Length 118;
 Best Local Similarity 51.3%; Pred. No. 7.5e-19;
 Matches 58; Conservative 17; Mismatches 34; Indels 4; Gaps 1;
 QY 54 RVRCSHLLVKHSQSRPSSWRQEK-----ITRTKEALELINGYIOIKSGEEDFESL 109
 Db 5 KVRASHLLKHQSGRRKRSKWDGDSLSATTRDDAVSQLSRLQDLPSPFSLASR 64
 QY 110 FDCSSAKARGDLGAFSGQMKPFEDASFAITGEMSGPVFTDSGIHILRT 162
 Db 65 HSHCSAKRGDGLGPGRGQMKPFEEATFALKVGESIDIVDTSVGHIIKRT 117
 RESULT 10
 QYXU8 PRELIMINARY; PRT; 123 AA.
 AC Q7XTU8 (T-EMBLrel. 25, Created)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE OSJNB0020P07.6 protein.
 GN OSJNB0020P07.6.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fu G., Wang S.Y., Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F.,
 RA Jia J., Yin H.F., Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y.,
 RA Shao Y., Sun Y., Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W.,
 RA Sheng H.H., Gu J.L., Chen S.T., Ni L., Zhu F.H., Han B., Feng Q.,
 RA Huang Y.C., Li Y., Zhu J.Q., Zhao Q., Hu X., Liu Y.L., Mu J., Yu Z.,
 RA Chen L., Fan D.L., Weng Q.J., Zhang L., Lu Y.Q., Yu S.L., Liu X.H.,
 RA Lu T.T., Zhang Y.J., Lu Y., Li C., Li T., Zhang Y., Hu H., Jia P.X.,
 RA Qian Y.M., Ying K., Zhou B., Chen Z.H., Hao P., Zhang L., Wu M.,
 RA Zhang R.Q., Guan J.P., Hong G.F.,
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL606450; CAB01289.1; -;
 SQ SEQUENCE 123 AA; 13340 MW; 511DAB2D4935E77 CRC64;
 Query Match 32.9%; Score 280; DB 10; Length 123;
 Best Local Similarity 50.9%; Pred. No. 1.5e-18;
 Matches 59; Conservative 15; Mismatches 38; Indels 4; Gaps 1;

QY 51 EPARVCSHLLVKHSQSRPSSWRQEK-----ITRTKEALELINGYIOIKSGEEDFESL 106
 Db 7 ETEIVRALHILIKHSGRRKRSKWDGDSLSATTRDDAVSQLSRLQDLPSPFSLASR 66
 QY 107 ASQFSDCSSAKARGDLGAFSGQMKPFEDASFAITGEMSGPVFTDSGIHILRT 162
 Db 67 AARHSDCSSARRGGDLGTGFRQMKPFEDATFALKVGESIDIVDTSVGHIIKRT 122
 RESULT 11
 QYXU8 PRELIMINARY; PRT; 386 AA.
 AC Q8IRJ5 (T-EMBLrel. 23, Created)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
 DE CG32845-PA.
 GN CG32845.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.R., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazei R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
 RA Abril J.F., Ayabayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Flesler C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nusskern D.R., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spiers E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Ye J.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferrera S., Frise E., Galie R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,

```

RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moehrefi A.,
RA Mcintosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskaas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RP Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
[5]
RN SEQUENCE FROM N.A.
RP FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003467; AANL1433.1; -.
DR FlyBase; FBgn0052845; CG32845.
DR InterPro; IPR000297; Rotamase.
DR InterPro; IPR001202; WW Rsp5_WWP.
DR Pfam; PF00639; Rotamase; 1.
DR Pfam; PF00397; WW; 1.
DR SMART; SM00456; WW; 1.
DR PROSITE; PS01096; PPIC_PPIASE 1; 1.
DR PROSITE; PS01098; PPIC_PPIASE 2; 1.
DR PROSITE; PS50020; WW DOMAIN 2; 1.
SQ SEQUENCE 396 AA; 44342 MW; 4B2AB19F137A249A CRC64;

Query Match 32.5%; Score 277; DB 5; Length 386;
Best Local Similarity 38.8%; Pred. No. 1.2e-17;
Matches 64; Conservative 27; Mismatches 66; Indels 8; Gaps 1;

Qy 6 KLPFGWEKRSRSGRVVYFNHITNASOWERPSPGNSGGKNGQGE-----PARVRC 57
Db 72 KLPFGWEERIAHSKCEVYFTIRKVFHTPPSHREKDRNANGALGDYSDFNQRLC 131
Qy 58 SHLVKHSQSRPSPRSRQEKITRTKEALELINGYIQIKSGBEDFESLASQFDCSSAK 117
Db 132 RHILVKHSESRCSSYRERWVRTKQKALNKIMHARDLIQSGKTFEAFELANNISDCCSAR 191
Qy 118 ARGDLGFSRQMKPFEDASFALRTGEMSGPVFTDSGIHILRT 162
Db 192 HGGDLGPLSLTQTPVFFERNILLKQGLSELFTQKAGYHILRT 236

RESULT 12
Q42334 PRELIMINARY; PRT; 69 AA.
ID Q42334
AC Q42334
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Peptidyl prolyl cis-trans isomerase (Fragment).
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Seed;
RA Cooke R., Laurie M., Raynal M., Delseny M.;


```

```

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; F13919; CAA23077.1; -.
DR HSSP; Q13526; 1PIN
GO; GO:0001685; F:isomerase activity; IEA.
DR InterPro; IPR000297; Rotamase.
DR Pfam; PF00639; Rotamase; 1.
DR PROSITE; PS01096; PPIC_PPIASE 1; 1.
DR PROSITE; PS01098; PPIC_PPIASE 2; 1.
KW Isomerase.
FT NON_TER 1
SQ SEQUENCE 69 AA; 7333 MW; 03C2DF735ED6F253 CRC64;

Query Match 25.7%; Score 219; DB 10; Length 69;
Best Local Similarity 62.7%; Pred. No. 3.9e-13;
Matches 42; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

Qy 96 IKSGBEDFESLASQFDCSSAKARGDLGAFSGQWQKPFEDASFALRTGEMSGPVFTDSG 155
Db 2 IVSGKANFEVATRVSDCSSAKRGDGLSGFGRGQWQKPFEEATYALKVGDISDIVDTDSG 61
Qy 156 IHIILRT 162
Db 62 VHIILRT 68

RESULT 13
Q8SR55 PRELIMINARY; PRT; 150 AA.
ID Q8SR55
AC Q8SR55;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Peptidyl prolyl cis trans isomerase (NIMA-interacting).
GN ECU06 0330.
OS Enecephalitozoon cuniculi.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Enecephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope;
RA Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=GB-M1;
RC MEDLINE=21576510; PubMed=11719806;
RX Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Frensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
Encephalitozoon cuniculi."
RL Nature 414:450-453(2001).
DR EMBL; AL590446; CAD25393.1; -.
DR InterPro; IPR000297; Rotamase.
DR Pfam; PF00639; Rotamase; 1.
DR PROSITE; PS01098; PPIC_PPIASE 2; 1.
SQ SEQUENCE 150 AA; 17203 MW; 51AA3C8864421487 CRC64;

Query Match 25.6%; Score 218.5; DB 5; Length 150;
Best Local Similarity 35.0%; Pred. No. 1.1e-12;
Matches 56; Conservative 24; Mismatches 53; Indels 27; Gaps 5;

Qy 8 PPGWEKMSRSGRVVYFNHITNASOWERPSPGNSGGKNGQGEPPARVRCSHLLVKSQS 67
Db 11 PEMVKKDKRGTGSPFYNTETAERTKPN-----EG----FRLYHILIKHKS 56
Qy 68 RPPSWRQEKITRTKEALELINGYIQ--KIKSGEDFESL-----ASQFSDCSSAKARGD 121
Db 57 RKP-----VDMSIDFAFSRIKAIHEDLRKAGDKNFKELFKDAIKHKSQCSAKRGD 109
Qy 122 LGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGIHILR 161

```

Db 110 LGFVCGNMMKEPEKPAFSLGRGEMSGVPSTPSGFHIIYR 149

RESULT 14

Q82SU8 PRELIMINARY; PRT; 630 AA.
AC Q82SU8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PiC-type peptidyl-prolyl cis-trans isomerase (EC 5.2.1.8).
GN PFID OR NE2206.
OS Nitrosomonas europaea.
OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
OC Nitrosomonadaceae; Nitrosomonas.
OX NCBI_TaxID=915;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19718 / IF0 14298;
RX MEDLINE=22586410; PubMed=12700255;
RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
HAUSER L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
RA Acierio D.M., Hommes N.G., Whittaker M.M., Arp D.J.;
RT "Complete genome sequence of the ammonia-oxidizing bacterium and
obligate chemolithoautotroph Nitrosomonas europaea.";
RL J. Bacteriol. 185:2759-2773 (2003).
DR EMBL; BX321863; CAD86118.1; -.
DR GO; GO:0042027; F:cyclophilin-type peptidyl-prolyl cis-trans . . .; IEA.
DR GO; GO:0004600; F:cyclophilin; IEA.
DR GO; GO:0030051; F:FK506-sensitive peptidyl-prolyl cis-trans i. . .; IEA.
DR GO; GO:0016853; F:isomerase activity; IEA.
DR InterPro; IPR000297; Rotamase.
DR Pfam; PF00639; Rotamase; 1.
DR PROSITE; PS0198; PPIC_PPIASE_2; 1.
KW Isomerase; Complete proteome.
SQ SEQUENCE 630 AA; 71638 MW; 35F9D1AFFFC34153 CRC64;

Query Match 21.7%; Score 185; DB 16; Length 630;

Best Local Similarity 32.7%; Pred. No. 8.5e-09;
Matches 50; Conservative 25; Mismatches 50; Indels 28; Gaps 4;

QY 12 EKRMSSGRVYFNFHITNASQWERPSSGGKNGGEPARVRCSHLLVKHSQSRPS 71
Db 241 QEQISDEIRKYDEH-----QDFGQAEERRASHILLS-----VPA 277
QY 72 SWRQKTRTKEALELINGYQIKSGEEDFESLASQFS-DCSSAKARGDLGAFSRGOM 130
Db 278 DATEQKTSKARAEQI-----LEQVRQDEKLPFLAALSELDPGSAKEGDLGFFARGLM 333
QY 131 OKPFEDASFALRTGEMSGPVFTDSGHIILRT 163
Db 334 VKPFEDVFMQGEIRGVPETPFPGFHIILRT 366

RESULT 15

Q899G4 PRELIMINARY; PRT; 246 AA.
AC Q899G4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative peptidyl-prolyl cis-trans isomerase.
GN CTC00215.
OS Clostridium tetani.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Massachusetts / E88;
RX MEDLINE=22457253; PubMed=12552129;
RA Brueggemann H., Baumer S., Fricke W.F., Wierer A., Liesegang H.,
RA Decker I., Herzberg C., Martinez-Arias R., Merkl R., Henne A.,

RA Gottschalk G.;
RT "The genome sequence of Clostridium tetani, the causative agent of
tetanus disease.";

RL Proc. Natl. Acad. Sci. U.S.A. 100:1316-1321 (2003).

DR EMBL; AE015936; AAC34863.1; -.

DR GO; GO:0016853; F:isomerase activity; IEA.

DR InterPro; IPR000297; Rotamase.

DR Pfam; PF00639; Rotamase; 1.

DR PROSITE; PS01096; PPIC_PPIASE_1; 1.

DR PROSITE; PS0198; PPIC_PPIASE_2; 1.

KW Isomerase; Complete proteome.

SQ SEQUENCE 246 AA; 28688 MW; C6A29D8AC52640A6 CRC64;

Query Match 21.2%; Score 181; DB 16; Length 246;

Best Local Similarity 38.5%; Pred. No. 6.5e-09;

Matches 45; Conservative 17; Mismatches 33; Indels 22; Gaps 3;

QY 43 SGGKNGGEPARVRCSHLLVKHSQSRPSRWROEKITRTKEALELINGYQIKSGEED 102

Db 103 NNNKMMFIEEERVSTKHILVE-----TKEEARNIVD-----EIKNG-LS 140

QY 103 FESLASQFDCSSAKARGDLGAFSRGOMQKPFEDASFALRTGEMSGPVFTDSGHIH 159

Db 141 FEEAAKEYSNCPKSGAGGDLGTPGGRWVKEFEAEAFEMKEGITSNPVKTFQFGYHII 197

Search completed: June 28, 2004, 18:39:08

Job time : 130 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2004, 14:35:24 ; Search time 4362.63 Seconds
(without alignments)
9875.458 Million cell updates/sec

Title: U49070
Perfect score: 994
Sequence: 1 TGTGGCCAGCACCTCGAGG.....CCCAATTAAACCCAGAACCA 994

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	994	100.0	994	6	AX214524	AX214524 Sequence
2	994	100.0	994	9	HSU49070	U49070 Human pepti
3	994	100.0	1014	6	AR074009	AR074009 Sequence
4	994	100.0	1014	6	AR082120	AR082120 Sequence
5	994	100.0	1014	6	AR360786	AR360786 Sequence
6	956	96.2	1021	9	BC002899	BC002899 Homo sapi
7	892.6	89.8	3125	6	AX747830	AX747830 Sequence
8	892.6	89.8	3125	9	AK092970	AK092970 Homo sapi
9	885.6	89.0	996	9	HSU82382	U82382 Homo sapien
10	885.6	89.0	153750	9	AL391728	AL391728 Human DNA
11	567.4	57.1	1656	6	AX834874	AX834874 Sequence
12	567.4	57.1	1656	9	AK097558	AK097558 Homo sapi
13	567.4	57.1	138036	9	AC008752	AC008752 Homo sapi
14	567.4	57.1	194624	9	AC008742	AC008742 Homo sapi
15	408.6	41.1	954	10	AB009691	AB009691 Mus sp. m
16	408.6	41.1	936	10	BC038254	BC038254 Mus muscu
17	375.2	37.7	138152	10	AL954170	AL954170 Mouse DNA
18	346.4	34.8	150017	2	AL773509	AL773509 Mus muscu
19	338	34.0	521	9	HUMZE03B04	AF086512 Homo sapi
20	321.2	32.3	164603	2	AC119466	AC119466 Rattus no
21	321.2	32.3	223996	2	AC094504	AC094504 Rattus no
22	319	32.1	529	5	AF239760	AF239760 Xenopus l
23	278.4	28.0	194880	2	AC091329	AC091329 Homo sapi
24	277.2	27.9	852	5	BC059553	BC059553 Danio rer
25	239.6	24.1	194880	2	AC091929	AC091929 Homo sapi
26	216	21.7	64036	2	AC040921	AC040921 Homo sapi
27	208.4	21.0	439	6	AX305342	AX305342 Sequence
28	161.4	16.2	993	3	DMU35140	U35140 Drosophila
29	159.8	16.1	17559	10	AB009692	AB009692 Mus sp. g
30	158.2	15.9	240584	2	AC073676	AC073676 Mus muscu
31	158.2	15.9	240584	2	AC116063	AC116063 Rattus no
32	158.2	15.9	241086	2	AC106653	AC106653 Rattus no
33	158.2	15.9	254452	2	AC118348	AC118348 Rattus no
34	158.2	15.9	255299	2	AC109676	AC109676 Rattus no
35	115.6	11.6	65675	2	AC026985	AC026985 Homo sapi
36	115.6	11.6	65675	2	AC026985	AC026985 Homo sapi
37	113.4	11.4	589	11	BV093798	BV093798 RPAMSEQ0
38	107.4	10.8	770	8	AY374446	AY374446 Paracocci
39	107.4	10.8	770	8	AY374446	AY374446 Pichia an
40	105.2	10.6	3415	8	AF350242	AF350242 Pichia an
41	100.6	10.1	607	11	BV093800	BV093800 RPAMSEQ0
42	99	10.0	66669	3	AF017777	AF017777 Drosophil
43	98.4	9.9	87089	2	AC013980	AC013980 Drosophil
44	98.4	9.9	171548	3	AC092494	AC092494 Drosophil
45	98.4	9.9	173508	3	AC011251	AC011251 Drosophil

ALIGNMENTS

RESULT 1
AX214524
LOCUS AX214524 994 bp DNA linear PAT 06-SEP-2001
DEFINITION Sequence 3 from Patent WO0159155.
ACCESSION AX214524
VERSION AX214524.1 GI:15524567
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Meritet J.F., Dron.M. and Tovey,M.G.
TITLE Interferon-alpha induced genes
JOURNAL Patent: WO 0159155-A 3 16-AUG-2001;

Pharma Pacific Pty. Ltd. (AU)
Location/Qualifiers
1. .994
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
25. .516
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC69362.1"
/db_xref="GI:15524568"
/db_xref="RETRMBL:CAC69362"
/translation="MADEEKLPGWEKMSRSGRVYVFNHITNASOWERPSGNSSSG
GKNGQGPAPRVCRSHLIVKHSQSRPSSWRQEKITRKEALELINGYIQIKSGEED
FESLASQFSDSSAKARGDLGAFSGMQKPFEDASFALTRGEMSGPVFTDSGIHIL
RTE"

ORIGIN

Query Match 100.0%; Score 994; DB 6; Length 994;
Best Local Similarity 100.0%; Pred. No. 1.4e-199; Indels 0; Gaps 0;
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCTGGCCAGCACCTCGAGGGAAGATGCGGACGAGGAGAGCTGCCGCCCGCTGGGAG 60
DB 1 TGCTGGCCAGCACCTCGAGGGAAGATGCGGACGAGGAGAGCTGCCGCCCGCTGGGAG 60
QY 61 AAGCGCATGAGCCGAGCTCAGGCGGAGTGTAATCTCAACACATCACTAAGCCGAGC 120
DB 61 AAGCGCATGAGCCGAGCTCAGGCGGAGTGTAATCTCAACACATCACTAAGCCGAGC 120
QY 121 CAGTGGGAGCGGCCAGCGGCAACAGCAGCAGTGGTGCAAAAACGGCGAGGGGAGCCT 180
DB 121 CAGTGGGAGCGGCCAGCGGCAACAGCAGCAGTGGTGCAAAAACGGCGAGGGGAGCCT 180
QY 181 GCCAGGCTCCGCTGCTGCACCTCTGCTGTAAGCAGACAGCTCAAGCGCGCCCTCGTCC 240
DB 181 GCCAGGCTCCGCTGCTGCACCTCTGCTGTAAGCAGACAGCTCAAGCGCGCCCTCGTCC 240
QY 241 TGGCGGAGGAGAGATCACCGGACCAAGGAGGAGGCGCTGGAGCTGATCAAGCGGTAC 300
DB 241 TGGCGGAGGAGAGATCACCGGACCAAGGAGGAGGCGCTGGAGCTGATCAAGCGGTAC 300
QY 301 ATCCAGAGATCAAGTCCGGAGAGGAGGACTTGGCTCTCGGCTCAGCTTCAGCGAC 360
DB 301 ATCCAGAGATCAAGTCCGGAGAGGAGGACTTGGCTCTCGGCTCAGCTTCAGCGAC 360
QY 361 TGCAGCTCAGCCAGCGGAGAGCTGGGTGCTTTCAGGAGGTCAGATCGAGAG 420
DB 361 TGCAGCTCAGCCAGCGGAGAGAGCTGGGTGCTTTCAGGAGGTCAGATCGAGAG 420
QY 421 CCATTGAGAGCGCTCGTTGCGCTGCGGAGCGGAGAGTGAAGCGGCGCTGTTACG 480
DB 421 CCATTGAGAGCGCTCGTTGCGCTGCGGAGCGGAGAGTGAAGCGGCGCTGTTACG 480
QY 481 GATTCGGGATCCATCATCTCCGCACTGAGTGGGAGGAGCGGAGCGGCTGGGCT 540
DB 481 GATTCGGGATCCATCATCTCCGCACTGAGTGGGAGGAGCGGAGCGGCTGGGCT 540
QY 541 CGGGGAGGAGGCGGCTAGGCGGCGAGCTCCCTTGCCTCCCGCAGCAGTGGCGGA 600
DB 541 CGGGGAGGAGGCGGCTAGGCGGCGAGCTCCCTTGCCTCCCGCAGCAGTGGCGGA 600
QY 601 ACCCCCCACTCCCTGCACCGTCACACAGTATTTATTTGTTCCCAATGGCTGGGAGGG 660
DB 601 ACCCCCCACTCCCTGCACCGTCACACAGTATTTATTTGTTCCCAATGGCTGGGAGGG 660
QY 661 GGCCTTCCAGATTGGGGGCGCTGGGTCCCACTCCCTGCTCCATCCCCAGTTGGGGTGC 720
DB 661 GGCCTTCCAGATTGGGGGCGCTGGGTCCCACTCCCTGCTCCATCCCCAGTTGGGGTGC 720
QY 721 GACCGGAGATTCCTTTAAGGAATTGACTTCAGCAGGAGGTTGGAGGCTCCAGACCCA 780
DB 721 GACCGGAGATTCCTTTAAGGAATTGACTTCAGCAGGAGGTTGGAGGCTCCAGACCCA 780

QY 781 GGGCAGTGTGGTGGGAGGGGTGTTCCAAAGAGAAGSCCTGGTCAGCAGAGCGGCCCGTG 840
DB 781 GGGCAGTGTGGTGGGAGGGGTGTTCCAAAGAGAAGSCCTGGTCAGCAGAGCGGCCCGTG 840
QY 841 TCCCCCCCAGGTGCTGGAGGAGCAGACTCGAGGCGCGAATTTCTTCTAGTTAGGCACTCC 900
DB 841 TCCCCCCCAGGTGCTGGAGGAGCAGACTCGAGGCGCGAATTTCTTCTAGTTAGGCACTCC 900
QY 901 TCTGTTCACTCGCAAGGTGAACACTCATGCGGAGGAGGCGCTCTGAGCAACTGTG 960
DB 901 TCTGTTCACTCGCAAGGTGAACACTCATGCGGAGGAGGCGCTCTGAGCAACTGTG 960
QY 961 CAGACCTTTTCAACCCCAATTAACCCAGAACCA 994
DB 961 CAGACCTTTTCAACCCCAATTAACCCAGAACCA 994

RESULT 2
HSU49070
LOCUS
DEFINITION
Human peptidyl-prolyl isomerase and essential mitotic regulator
(PIN1) mRNA, complete cds.
ACCESSION
U49070
VERSION
U49070.1
KEYWORDS
peptidyl-prolyl isomerase; cell cycle; mitotic regulator; NIMA;
Eas1; NIMA-interacting protein 1 (Pin1).
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 994)
AUTHORS
Lu, K.P., Hanes, S.D. and Hunter, T.
TITLE
A human peptidyl-prolyl isomerase essential for regulation of
mitosis 380 (5574), 544-547 (1996)
JOURNAL
NATURE 380 (5574), 544-547 (1996)
MEDLINE
96195064
PUBMED
8606777
REFERENCE
2 (bases 1 to 994)
AUTHORS
Lu, K.P. and Hunter, T.
TITLE
Direct Submission
JOURNAL
Submitted (12-FEB-1996) Kun Ping Lu, Molecular Biology and Virology
Laboratory, Salk Institute, 10010 North Torrey Pines Rd., La Jolla,
CA 92037, USA
FEATURES
source
1. .994
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="HeLa"
gene
1. .994
/gene="PIN1"
CDS
25. .516
/gene="PIN1"
/note="NIMA-interacting protein 1, essential mitotic
regulator, essential peptidyl-prolyl isomerase"
/codon_start=1
/product="pin1"
/protein_id="AAC50492.1"
/db_xref="GI:1332710"
/translation="MADEEKLPGWEKMSRSGRVYVFNHITNASOWERPSGNSSSG
GKNGQGPAPRVCRSHLIVKHSQSRPSSWRQEKITRKEALELINGYIQIKSGEED
FESLASQFSDSSAKARGDLGAFSGMQKPFEDASFALTRGEMSGPVFTDSGIHIL
RTE"

ORIGIN

Query Match 100.0%; Score 994; DB 9; Length 994;
Best Local Similarity 100.0%; Pred. No. 1.4e-199;
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCTGGCCAGCACCTCGAGGGAAGATGCGGACGAGGAGAGCTGCCGCCCGCTGGGAG 60
DB 1 TGCTGGCCAGCACCTCGAGGGAAGATGCGGACGAGGAGAGCTGCCGCCCGCTGGGAG 60


```

QY      841  TCCCTCCAGGTCCTGGAGCAGACTCGAGGCGCCGAATTTCTAGTTAGGCCACGCTCC 900
Db      841  TCCCTCCAGGTCCTGGAGCAGACTCGAGGCGCCGAATTTCTAGTTAGGCCACGCTCC 900
QY      901  TCTGTTCAGTCGCAAGGTGAACACTCATGCGGCGACCATGGCCCTCTGAGCAACTGTG 960
Db      901  TCTGTTCAGTCGCAAGGTGAACACTCATGCGGCGACCATGGCCCTCTGAGCAACTGTG 960
QY      961  CAGACCCCTTTCACCCCAATTAACCCAGAACCA 994
Db      961  CAGACCCCTTTCACCCCAATTAACCCAGAACCA 994

RESULT 4
LOCUS      AR082120      1014 bp      DNA      linear      PAT 31-AUG-2000
DEFINITION Sequence 1 from patent US 5972697.
ACCESSION  AR082120
VERSION     AR082120.1  GI:10008846
KEYWORDS   .
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 1014)
AUTHORS     Hunter,T. and Lu,K.Ping.
TITLE       NIMA interacting proteins
JOURNAL     Patent: US 5972697-A 1 26-OCT-1999;
FEATURES    Location/Qualifiers
             source
             1..1014
             /organism="unknown"
             /mol_type="unassigned DNA"

ORIGIN
Query Match      100.0%; Score 994; DB 6; Length 1014;
Best Local Similarity 100.0%; Pred. No. 1.4e-199;
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  TGTGGCCAGCACCTCAGGAGAGTGGCGGACGAGGAGAGTGGCCCGGCTGGGAG 60
Db      1  TGTGGCCAGCACCTCAGGAGAGTGGCGGACGAGGAGAGTGGCCCGGCTGGGAG 60
QY      61  AAGGCGATGAGCCGAGTCAACCGGACCAAGGAGAGGCGCTGGAGCTGATCAACGGCTAC 120
Db      61  AAGGCGATGAGCCGAGTCAACCGGACCAAGGAGAGGCGCTGGAGCTGATCAACGGCTAC 120
QY      121  CAGTGGGAGCGGCCAGCGGCAACAGCAGCAGTGGTGGCAAAAACGGGCGAGGAGCT 180
Db      121  CAGTGGGAGCGGCCAGCGGCAACAGCAGCAGTGGTGGCAAAAACGGGCGAGGAGCT 180
QY      181  GCCAGGTCCTGCTGCTGCACTGCTGCTGAGCAGCAGCCAGTCAAGGCGGCTCGTCC 240
Db      181  GCCAGGTCCTGCTGCTGCACTGCTGCTGAGCAGCAGCCAGTCAAGGCGGCTCGTCC 240
QY      241  TGGCGGAGGAGAGTCAACCGGACCAAGGAGAGGCGCTGGAGCTGATCAACGGCTAC 300
Db      241  TGGCGGAGGAGAGTCAACCGGACCAAGGAGAGGCGCTGGAGCTGATCAACGGCTAC 300
QY      301  ATCCAGAGATCAAGTCGGGAGAGAGGACTTTGAGTCTCTGCGCTCAAGTTCAAGCAG 360
Db      301  ATCCAGAGATCAAGTCGGGAGAGAGGACTTTGAGTCTCTGCGCTCAAGTTCAAGCAG 360
QY      361  TGCAGTCAAGCAGGCGAGGAGACCTGGGTCCTTCAGCAGAGGTGATGATGAGAG 420
Db      361  TGCAGTCAAGCAGGCGAGGAGACCTGGGTCCTTCAGCAGAGGTGATGATGAGAG 420
QY      421  CCATTTGAAGACGCTTCGTTTTCGCTGCGGACGGGGAGATGAGCGGGCCCGTTTACG 480
Db      421  CCATTTGAAGACGCTTCGTTTTCGCTGCGGACGGGGAGATGAGCGGGCCCGTTTACG 480
QY      481  GATTCGGCATCCACATCATCTCCGCACTGAGTGGGAGCCCGGCTGGCT 540
Db      481  GATTCGGCATCCACATCATCTCCGCACTGAGTGGGAGCCCGGCTGGCT 540

```

```

QY      541  CGGGCAGGGCAGGGCGGCTAGGCGGCGCAGCTCCCTTTGCCGCCAGCCAGTGGCCGA 600
Db      541  CGGGCAGGGCAGGGCGGCTAGGCGGCGCAGCTCCCTTTGCCGCCAGCCAGTGGCCGA 600
QY      601  ACCCCCACTCCCTGCGCACCGTCAACAGTATTTATTTGTTCCCAATGCTGGAGGG 660
Db      601  ACCCCCACTCCCTGCGCACCGTCAACAGTATTTATTTGTTCCCAATGCTGGAGGG 660
QY      661  GCGCTTCAGATTGGGGCCCTGGGGTCCCACTCCCTGTCCATCCCACTGGGGTGC 720
Db      661  GCGCTTCAGATTGGGGCCCTGGGGTCCCACTCCCTGTCCATCCCACTGGGGTGC 720
QY      721  GACCGCCAGATTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGAGGCTCCAGACCA 780
Db      721  GACCGCCAGATTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGAGGCTCCAGACCA 780
QY      781  GGGCAGTGTGGGAGGGGTGTTCCAAAGAGAGGCGCTGTCAGCAGAGCGGCCCGTG 840
Db      781  GGGCAGTGTGGGAGGGGTGTTCCAAAGAGAGGCGCTGTCAGCAGAGCGGCCCGTG 840
QY      841  TCCCGCCAGGTGCTGAGGCGAGCTCGAGGGCGGAATTTCTAGTTAGGCCAGCTCC 900
Db      841  TCCCGCCAGGTGCTGAGGCGAGCTCGAGGGCGGAATTTCTAGTTAGGCCAGCTCC 900
QY      901  TCTGTTCAGTCGCAAGGTGAACACTCATGCGGCGACCATGGCCCTCTGAGCAACTGTG 960
Db      901  TCTGTTCAGTCGCAAGGTGAACACTCATGCGGCGACCATGGCCCTCTGAGCAACTGTG 960
QY      961  CAGACCCCTTTCACCCCAATTAACCCAGAACCA 994
Db      961  CAGACCCCTTTCACCCCAATTAACCCAGAACCA 994

RESULT 5
LOCUS      AR360786      1014 bp      DNA      linear      PAT 17-AUG-2003
DEFINITION Sequence 1 from patent US 6596848.
ACCESSION  AR360786
VERSION     AR360786.1  GI:33768227
KEYWORDS   .
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 1014)
AUTHORS     Hunter,T. and Lu,K.P.
TITLE       Antibodies to NIMA interacting proteins
JOURNAL     Patent: US 6596848-A 1 22-JUL-2003;
FEATURES    Location/Qualifiers
             source
             1..1014
             /organism="unknown"
             /mol_type="genomic DNA"

ORIGIN
Query Match      100.0%; Score 994; DB 6; Length 1014;
Best Local Similarity 100.0%; Pred. No. 1.4e-199;
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  TGTGGCCAGCACCTCAGGAGGAGATGGCGGACGAGGAGAGTGGCCCGGCTGGGAG 60
Db      1  TGTGGCCAGCACCTCAGGAGGAGATGGCGGACGAGGAGAGTGGCCCGGCTGGGAG 60
QY      61  AAGGCGATGAGCCGAGTCAACCGGACCAAGGAGAGGCGCTGGAGCTGATCAACGGCTAC 120
Db      61  AAGGCGATGAGCCGAGTCAACCGGACCAAGGAGAGGCGCTGGAGCTGATCAACGGCTAC 120
QY      121  CAGTGGGAGCGGCCAGCGGCAACAGCAGCAGTGGTGGCAAAAACGGGCGAGGAGCT 180
Db      121  CAGTGGGAGCGGCCAGCGGCAACAGCAGCAGTGGTGGCAAAAACGGGCGAGGAGCT 180
QY      181  GCCAGGTCCTGCTGCTGCACTGCTGCTGAGCAGCAGCCAGTCAAGGCGGCTCGTCC 240
Db      181  GCCAGGTCCTGCTGCTGCACTGCTGCTGAGCAGCAGCCAGTCAAGGCGGCTCGTCC 240

```


241	QY	TGGGGCAGGAGAGTACACCGGACCAAGGAGGAGGCGCTGGAGCTGATCAACGGCTAC	300
241	Db	TGGGGCAGGAGAGTACACCGGACCAAGGAGGAGGCGCTGGAGCTGATCAACGGCTAC	300
301	QY	ATCCAGAGATCAAGTCGGGAGAGGAGGACTTTGAGTCTCTGCGCTCACAGTTTCAGCGAC	360
301	Db	ATCCAGAGATCAAGTCGGGAGAGGAGGACTTTGAGTCTCTGCGCTCACAGTTTCAGCGAC	360
361	QY	TGCAGCTCAGCCCAAGGCGAGGAGAGCTGGGTGCTTCAGCGAGGTTCAGATGCGAGAAG	420
361	Db	TGCAGCTCAGCCCAAGGCGAGGAGAGCTGGGTGCTTCAGCGAGGTTCAGATGCGAGAAG	420
421	QY	CCATTTGAAGACGCTGTTTGCGCTGCGACGCGGAGAGATGACGGCGCGCTGTTTCACG	480
421	Db	CCATTTGAAGACGCTGTTTGCGCTGCGACGCGGAGAGATGACGGCGCGCTGTTTCACG	480
481	QY	GATTCGGGCTATCCATCATCTCCGCACTGAGTGAGGGTGGGAGCCCGGCTGCGCT	540
481	Db	GATTCGGGCTATCCATCATCTCCGCACTGAGTGAGGGTGGGAGCCCGGCTGCGCT	540
541	QY	CGGGCGAGGCGAGGCGGCTAGGCGGCGAGCTCCCTTCCCGCCAGCAGTGGCGGA	600
541	Db	CGGGCGAGGCGAGGCGGCTAGGCGGCGAGCTCCCTTCCCGCCAGCAGTGGCGGA	600
601	QY	ACCCGCCACTCCCTGCCACGTCACAGATATTTATTGTTCCCAATGGCTGGGAGGGG	660
601	Db	ACCCGCCACTCCCTGCCACGTCACAGATATTTATTGTTCCCAATGGCTGGGAGGGG	660
661	QY	GCCCTTCCAGATTTGGGGGCTGGGGTCCCGACCTCCCTGTCATCCCGAGTGGGGTGC	720
661	Db	GCCCTTCCAGATTTGGGGGCTGGGGTCCCGACCTCCCTGTCATCCCGAGTGGGGTGC	720
721	QY	GACCGCCAGATTTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGGAGCTCCCGACCCA	780
721	Db	GACCGCCAGATTTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGGAGCTCCCGACCCA	780
781	QY	GGGCGAGTGTGGGAGGGGTGTTCCAAAGAGAGCGCTGGTCAGCAGCGCGCCCGCTG	840
781	Db	GGGCGAGTGTGGGAGGGGTGTTCCAAAGAGAGCGCTGGTCAGCAGCGCGCCCGCTG	840
841	QY	TCCCCCGAGGTGCTGGAGGCGAGACTCGAGGGCGGAAATGTTTCTAGTTAGGCGCACTCC	900
841	Db	TCCCCCGAGGTGCTGGAGGCGAGACTCGAGGGCGGAAATGTTTCTAGTTAGGCGCACTCC	900
901	QY	TCTGTTTCAGTCCAAAGGTGAACATCATGCGGCGAGCCATGGCGCTCTGAGCAACTGTG	960
901	Db	TCTGTTTCAGTCCAAAGGTGAACATCATGCGGCGAGCCATGGCGCTCTGAGCAACTGTG	960
961	QY	CAGACCTTTTACCCCGCAATTAACCCAGAACCA	994
961	Db	CAGACCTTTTACCCCGCAATTAACCCAGAACCA	994
RESULT	6		
LOCUS	BC002899	1021 bp mRNA linear	PRI 12-NOV-2003
DEFINITION	Homo sapiens protein (peptidyl-prolyl cis/trans isomerase)		
	NIMA-interacting 1, mRNA (cDNA clone MGC:10717 IMAGE:3941595), complete cds.		
ACCESSION	BC002899	GI:12804092	
VERSION			
KEYWORDS	MGC.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 1021)		
AUTHORS	Strausberg, R.L., Feigold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenn, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullaby, S.J., Bosak, S.A., McEwan, P.J., McKernan, R.J., Malek, J.A., Gunaratne, P.H., Richards, S., Winkley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, B., Kettman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalins, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A. full-length human and mouse cDNA sequences		
	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE	2 (bases 1 to 1021)		
AUTHORS	Strausberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cgapbs-remail.nih.gov Tissue Procurement: DCTD/DTT cDNA Library Preparation: Rubin Laboratory DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (ILNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc_mgc@hgrl.nih.gov Ahter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Gupta, J., Haghighi, P., Dietrich, N.L., Granite, S., Guan, X., Li, J., Lian, J., Lian, J., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgenev, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, D.-H. and Green, E.D.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov Series: IRAL plate: 15 Row: m Column: 9 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5453897.		
FEATURES	Location/Qualifiers		
source	1..1021		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="MGC:10717 IMAGE:3941595"		
	/tissue_type="Lung, small cell carcinoma"		
	/clone_lib="NIH MGC 7"		
	/lab_host="DH10B-R"		
	/note="Vector: pOTB7"		
gene	1..1021		
	/gene="PIN1"		
	/note="synonym: DOD"		
	/db_xref="LocusID:5300"		
	/db_xref="MIM:601052"		
	28..519		
	/codon_start=1		
	/product="protein (peptidyl-prolyl cis/trans isomerase)"		
	NIMA-interacting 1"		
	/protein_id="AAH02899.1"		
	/db_xref="GI:12804093"		
	/db_xref="LocusID:5300"		
	/translation="MADEKLPQWEKMSRSGRVYFNHITNASQWEPNSGSSG		
	GNQGEPAVRCSLLLVKHSQSRRTSSRQEKITTKKEALELINGYIKIKSGED		
CDS			

FEELASQFSDCSSAKARGLGAFSRGQMQKPFEDASPALRTGEMSGPVFTDSGHILL
RTE"
46.138
/note="WW; Region: WW domain. The WW domain is a protein
module with two highly conserved tryptophans that binds
proline-rich peptide motifs in vitro"
/db_xref="CDD:pfam00397"
181.516
misc_feature
/note="Rotamase; Region: PFIC-type PPIASE domain.
Rotamases increase the rate of protein folding by
catalysing the interconversion of cis-proline and
trans-proline"
/db_xref="CDD:pfam00639"

ORIGIN

Query Match 96.2%; Score 956; DB 9; Length 1021;
Best Local Similarity 99.2%; Pred. No. 1.5e-191;
Matches 982; Conservative 0; Mismatches 5; Indels 3; Gaps 2;
QY 8 CAGCACCTCGAGGAGATGCGGACGAGGAGAGCTGCGCCGCGCTGGGAGAGCGCA 67
DB 11 CTGCGCAGGAGGAGATGCGGACGAGGAGAGCTGCGCCGCGCTGGGAGAGCGCA 70
QY 68 TGAGCGCAGCTCAGCGCGAGGTACTCTCAACACATCACTAAGCCAGCCAGTGGG 127
DB 71 TGAGCGCAGCTCAGCGCGAGGTACTCTCAACACATCACTAAGCCAGCCAGTGGG 130
QY 128 AGCGGCCAGCGCAACAGCAGCAGTGTGGCAAAACCGGCGAGGGGAGCTGCCAGGG 187
DB 131 AGCGGCCAGCGCAACAGCAGCAGTGTGGCAAAACCGGCGAGGGGAGCTGCCAGGG 190
QY 188 TCGCTGTGCACTCTGTGTGAGCAGCAGCAGTCAAGCGGCGCTGTCTGCGCGC 247
DB 191 TCGCTGTGCACTCTGTGTGAGCAGCAGCAGTCAAGCGGCGCTGTCTGCGCGC 250
QY 248 AGAGAGATCAACCGGACCAAGAGAGGCGCTGAGCTGATCAACGGCTACATCCAGA 307
DB 251 AGAGAGATCAACCGGACCAAGAGAGGCGCTGAGCTGATCAACGGCTACATCCAGA 310
QY 308 AGATCAAGTCGGGAGAGGAGCTTTGAGTCTTGGCTCAGAGTCAGGAGCTGAGCT 367
DB 311 AGATCAAGTCGGGAGAGGAGCTTTGAGTCTTGGCTCAGAGTCAGGAGCTGAGCT 370
QY 368 CAGCAGGCGGAGGAGCTGGTCTTACAGAGGAGGCGCTGAGCTGATCAACGGCTAC 427
DB 371 CAGCAGGCGGAGGAGCTGGTCTTACAGAGGAGGCGCTGAGCTGATCAACGGCTAC 430
QY 428 AAGAGCGCTCGTTGCGCTGCGGACGCGGAGAGTACGCGGCGCTTTCACGGATTCG 487
DB 431 AAGAGCGCTCGTTGCGCTGCGGACGCGGAGAGTACGCGGCGCTTTCACGGATTCG 490
QY 488 GCATCCACATCCTCCGCACTGAGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 547
DB 491 GCATCCACATCCTCCGCACTGAGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 550
QY 548 GGGCAGGCGGCTAGGCGGCGAGCTCCCTTCCCGCAGGAGGAGGAGGAGGAGGAGG 607
DB 551 GGGCAGGCGGCTAGGCGGCGAGCTCCCTTCCCGCAGGAGGAGGAGGAGGAGGAGG 610
QY 608 ACTCCCTCCACCGTCAACAGTATTTATTTGTCACAAATGGTGGAGGGGCGCTTC 667
DB 611 ACTCCCTCCACCGTCAACAGTATTTATTTGTCACAAATGGTGGAGGGGCGCTTC 670
QY 668 CAGATTGGGGGCGCTGGGGTCCCACTCCCTGTCCATCCCAAGTGGGGCTGGACCGCC 727
DB 671 CAGATTGGGGGCGCTGGGGTCCCACTCCCTGTCCATCCCAAGTGGGGCTGGACCGCC 730
QY 728 AGATTCTCCCTTAAGGAATTAAGTCTCAGCAGGGGTGGAGGCTCCCAAGACCCAGGCGAGT 787
DB 731 AGATTCTCCCTTAAGGAATTAAGTCTCAGCAGGGGTGGAGGCTCCCAAGACCCAGGCGAGT 790
QY 788 GTGGTGGGAGGGGTGTCCAAAGAGAGGCGCTGGTCAAGAGCGCGCGCGCTGTCCCGCC 847

DB 791 GTGGTGGGAGGGTGTTCAAAGAGAGAGGCGCTGGTCAAGAGAGCGCGCGCGCTGTCCCGCC 850
QY 848 AGGTGCTGGAGGAGGAGCTCGAGGCGCGAATTTCTTAGTTAGGCGCACGCTCTCTGTTTC 907
DB 851 AGGTGCTGGAGGAGGAGCTCGAGGCGCGAATTTCTTAGTTAGGCGCACGCTCTCTGTTTC 910
QY 908 AGTCGCAAGGTGAACACTCATGCGG--CAGCCATGCGGCGCTCTGAGCAACTGTGCGAG-A 964
DB 911 AGTCGCAAGGTGAACACTCATGCGGCGCGAGGCGCGCTCTGAGCAACTGTGCGAGCA 970
QY 965 CCCTTTACCGCCGAATTAACCCAGACCA 994
DB 971 CCCTTTACCGCCGAATTAACCCAGACCA 1000
RESULT 7
AX747830 3125 bp mRNA linear PAT 20-JUN-2003
LOCUS
DEFINITION Sequence 1355 from Patent EP1308459.
ACCESSION AX747830
VERSION AX747830.1 GI:32132218
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Iisagi, T., Sugiyama, T., Otsuka, T., Makamatsu, A., Sato, H., Ishii, S.,
Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,
Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and
Masuho, Y.
Full-length cDNA sequences
Patent: EP 1308459-A 1355 07-MAY-2003;
JOURNAL Helix Research Institute (JP); Research Association for
Biotechnology (JP)
FEATURES
source
1..3125
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 89.8%; Score 892.6; DB 6; Length 3125;
Best Local Similarity 98.7%; Pred. No. 3.3e-178;
Matches 921; Conservative 0; Mismatches 9; Indels 3; Gaps 2;
QY 65 GCATGAGCGCAGCTCAGGCGGAGTGTACTTCAACCACTACTTCAACCACTACTAAGCGGCGAGT 124
DB 2191 GCCTGCCCTCCCTCAGGCGGAGTGTACTTCAACCACTACTAAGCGGCGAGT 2250
QY 125 GGGAGCGCGCCAGCGGCAACAGCAGCAGTGTGGCAAAACGGGCGAGGGAGCGCTGCCA 184
DB 2251 GGGAGCGCGCCAGCGGCAACAGCAGCAGTGTGGCAAAACGGGCGAGGGAGCGCTGCCA 2310
QY 185 GGTTCGCTGCTCGACCTCTGTGAGCAGCAGCAGTCAAGCGGCGCTCTGCTGCTGGC 244
DB 2311 GGTTCGCTGCTCGACCTCTGTGAGCAGCAGCAGTCAAGCGGCGCTCTGCTGCTGGC 2370
QY 245 GGCAGGAGAGATCACCCGACCAAGAGGAGGCGCTTGGAGCTGATCAACGGGTACATCC 304
DB 2371 GGCAGGAGAGATCACCCGACCAAGAGGAGGCGCTTGGAGCTGATCAACGGGTACATCC 2430
QY 305 AGAAGATCAAGTCGGGAGAGAGGACTTTGAGTCTCTGGCTTCAAGTTCAGGAGCTGCA 364
DB 2431 AGAAGATCAAGTCGGGAGAGAGGACTTTGAGTCTCTGGCTTCAAGTTCAGGAGCTGCA 2490
QY 365 GCTCAGCCAGGCGCAGGAGAGCTGGTGTCTTACAGCAGAGTTCAGTTCAGAGGCGCAT 424
DB 2491 GCTCAGCCAGGCGCAGGAGAGCTGGTGTCTTACAGCAGAGTTCAGTTCAGAGGCGCAT 2550
QY 425 TTGAAGACGCTCTGCTGCTGCGCTGCGGACGGGAGAGTGAAGCGGCGCTGTTCACGAGT 484
DB 2551 TTGAAGACGCTCTGCTGCTGCGCTGCGGACGGGAGAGTGAAGCGGCGCTGTTCACGAGT 2610

Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES		source	
1. 3125		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="SPLEN2013670"	
		/tissue type="spleen"	
		/clone_lib="SPLEN2"	
		/note="cloning vector: pME18SFL3"	
ORIGIN			
Query Match		89.8%; Score 892.6; DB 9; Length 3125;	
Best Local Similarity		98.7%; Pred. No. 3.3e-178;	
Matches		921; Conservative 0; Mismatches 9; Indels 3; Gaps 2;	
QY	65	GCATGAGCCGACGCTCAGGCCGAGTGTACTTACACACATCATAACGCCAGCCAGT	124
Db	2191	GCCTGCCCTCCCTCCAGGCCGAGTGTACTTACACACATCATAACGCCAGCCAGT	2250
QY	125	GGGAGCGGCCGCGCAACAGCAGCAGTGTGTGGCAAAAACGGGCGAGGGGAGCTGCCA	184
Db	2251	GGGAGCGGCCGCGCAACAGCAGCAGTGTGTGGCAAAAACGGGCGAGGGGAGCTGCCA	2310
QY	185	GGGTCCGCTGCTCGACCTGCTGGTGAAGCAGCAGCAGTGTGTGGCAAAAACGGGCGAGGGGAGCTGCCA	244
Db	2311	GGGTCCGCTGCTCGACCTGCTGGTGAAGCAGCAGCAGTGTGTGGCAAAAACGGGCGAGGGGAGCTGCCA	2370
QY	245	GGCAGGAGAGATCACCGGACCAAGGAGGAGGCGCTGGAGCTGATCAACGGCTACATCC	304
Db	2371	GGCAGGAGAGATCACCGGACCAAGGAGGAGGCGCTGGAGCTGATCAACGGCTACATCC	2430
QY	305	AGAAAGATCAAGTCGGGAGAGAGGACTTTGAGTCTCTGGCTCACAGTTACGCACTGCA	364
Db	2431	AGAAAGATCAAGTCGGGAGAGAGGACTTTGAGTCTCTGGCTCACAGTTACGCACTGCA	2490
QY	365	GCTCAGCAAGCCGAGGGAGACCTGGTGTCTTTCAGCAGAGGTTCAGTGCAGAGCCAT	424
Db	2491	GCTCAGCAAGCCGAGGGAGACCTGGTGTCTTTCAGCAGAGGTTCAGTGCAGAGCCAT	2550
QY	425	TTGAAGAGCGCTCTTTCGCTGGGAGAGAGGAGGCGGCTGGTGTTCACGGATT	484
Db	2551	TTGAAGAGCGCTCTTTCGCTGGGAGAGAGGAGGCGGCTGGTGTTCACGGATT	2610
QY	485	CCGGCATCCACATCATCTCCGCACTGAGTGAAGGTGGGAGCCGAGGCTCGGCTCGG	544
Db	2611	CCGGCATCCACATCATCTCCGCACTGAGTGAAGGTGGGAGCCGAGGCTCGGCTCGG	2670
QY	545	GCAGGCGAGGGGCTAGCGGCGGCTCCCTTCCGCGGAGGAGGAGGAGGAGGAGGAGG	604
Db	2671	GCAGGCGAGGGGCTAGCGGCGGCTCCCTTCCGCGGAGGAGGAGGAGGAGGAGGAGG	2730
QY	605	CCCACTCCCTCCGACCGCTCACACAGTATTATTGTTTCCCAAAATGGCTGGAGGGGGCC	664
Db	2731	CCCACTCCCTCCGACCGCTCACACAGTATTATTGTTTCCCAAAATGGCTGGAGGGGGCC	2790
QY	665	TTCCAGATTGGGGGCGCTGGGCTCCCACTCCCTGTCATCCCAATGCTGGGCTGGGAGC	724
Db	2791	TTCCAGATTGGGGGCGCTGGGCTCCCACTCCCTGTCATCCCAATGCTGGGCTGGGAGC	2850
QY	725	GCCAGATTCTCCCTTAAGGAATTGACTTTCAGCAGGGGTGGAGGCTCCAGAGCCAGGGC	784
Db	2851	GCCAGATTCTCCCTTAAGGAATTGACTTTCAGCAGGGGTGGAGGCTCCAGAGCCAGGGC	2910
QY	785	AGTGTGTGGGAGGGGTGTTCCAAAGAGAGGCGCTGTGTCAAGAGGCGCCCGCTGTCC	844
Db	2911	AGTGTGTGGGAGGGGTGTTCCAAAGAGAGGCGCTGTGTCAAGAGGCGCCCGCTGTCC	2970
QY	845	CCCAAGTGTGGGAGGAGCTCGAGGGCGGAAATTTGTTAGTTAGGCGACGCTCTCTG	904
Db	2971	CCCAAGTGTGGGAGGAGCTCGAGGGCGGAAATTTGTTAGTTAGGCGACGCTCTCTG	3030
QY	905	TTCAAGTGTGGGAGGAGCTCGAGGGCGGAAATTTGTTAGTTAGGCGACGCTCTCTG	962
Db	3031	TTCAAGTGTGGGAGGAGCTCGAGGGCGGAAATTTGTTAGTTAGGCGACGCTCTCTG	3090
QY	963	G-ACCCTTTCACCCCAATTAAACCCAGAACCA 994	
Db	3091	GCACCCCTTTCACCCCAATTAAACCCAGAACCA 3123	
RESULT 8			
AK092970			
LOCUS			
Homo sapiens cdna FLJ35651 fis, clone SPLEN2013670.			
DEFINITION			
AK092970.1 GI:21751686			
VERSION			
oligo capping; fis (full insert sequence).			
KEYWORDS			
Homo sapiens (human)			
SOURCE			
Homo sapiens			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
1			
Ishibashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S.,			
Ono, Y., Hotuta, T., Hiraoka, S., Murakawa, K., Takiguchi, S.,			
Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M.,			
Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Sato, H.,			
Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,			
Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K.,			
Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M.,			
Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y.,			
Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isogai, T.			
NEDO human cDNA sequencing project			
Unpublished			
2 (bases 1 to 3125)			
Isogai, T. and Yamamoto, J.			
Direct Submission			
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7			
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan			
Submitted (04-JUL-2002) Tel: 81-438-52-3975, Fax: 81-438-52-3986)			
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)			
NEDO human cDNA sequencing project supported by Ministry of			
Economy, Trade and Industry of Japan; CDNA full insert sequencing:			
Research Association for Biotechnology (RAB); CDNA library			
construction: Helix Research Institute (HRI) (supported by Japan			
Key Technology Center etc.); 5' - & 3' - end one pass sequencing: RAB,			
HRI, and Biotechnology Center, National Institute of Technology and			

Tue Jun 29 08:37:21 2004

```

QY 905 TTCAAGTGCAGAGGTGAACACTCATGCGG--CAGCCATGGGGCCCTCTGAGCAACTGTGCA 962
Db 3031 TTCAAGTGCAGAGGTGAACACTCATGCGGCGCCAGCCATGGGGCCCTCTGAGCAACTGTGCA 3090
QY 963 G-ACCTTTACCCGCCCAATTAAACCCGAAACCA 994
Db 3091 GCACCCCTTACCCGCCCAATTAAACCCGAAACCA 3123

RESULT 9
HSU82382 996 bp mRNA linear PRI 21-OCT-1997
LOCUS Homo sapiens PIN1 peptidyl-prolyl cis/trans isomerase-like (PIN1L)
DEFINITION mRNA, complete cds.
ACCESSION U82382
VERSION U82382.1 GI:2501784
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 996)
Campbell, H.D., Webb, G.C., Fountain, S. and Young, I.G.
The human PIN1 peptidyl-prolyl cis/trans isomerase gene maps to
human chromosome 19p13 and the closely related PIN1L gene to 1p31
Genomics 44 (2), 157-162 (1997)
JOURNAL
MEDLINE 97446133
PUBMED 9299231
REFERENCE 2 (bases 1 to 996)
AUTHORS Campbell, H.D.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1996) Centre for Molecular Structure and
Function, RSBs, Australian National University, PO Box 475,
Canberra ACT 2601, Australia
FEATURES
source 1..996
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="19"
/map="19p13"
1..996
/gene="PIN1L"
36..338
/genes="PIN1L"
/notes="contains WW domain related to PIN1"
/codon_start=1
/product="PIN1 peptidyl-prolyl cis/trans isomerase-like"
/protein_id="AAB81333.1"
/db_xref="GI:2501785"
/translation="MADEKLPQWEKMSRPSRGVYFNHITNPSPWRPSGNSSG
KTIWQGEPAVRVRSLLVKKPKAALDIAAGHPDQGGPGADQLHPDQGRGL"
979..984
/polyA_signal /gene="PIN1L"
/evidence=experimental
996
/polyA_site /gene="PIN1L"
/evidence=experimental
994
ORIGIN
Query Match 69.0%; Score 685.6; DB 9; Length 996;
Best Local Similarity 86.1%; Pred. No. 1.9e-134;
Matches 861; Conservative 0; Mismatches 84; Indels 55; Gaps 7;
QY 17 GAGGGAAGATGCGGACGAGAGAGCTGCGCCCGCTGGAGAGCGCATGAGCCGCA 76
Db 28 GAGGGAAGATGCGGACGAGAGAACTCGCCCTGGTGGGAAAGCGCATGAGCCGCC 87
QY 77 GCTAGGCGGAGTGTACTTCAACCATCTCAACGCCAGCCAGTGGGAGCGGCCCA 136
Db 88 CCTCAGGCGGAGGTACTTCAACCATCTCAACGCCAGCCAGTGGGAGCGGCCCA 147
QY 137 GCGGCAACAGCAGCAGTGTGTGGCAAAACCGGCGAGGGGAGCCTCCGAGGGTCCGCTGT 196

```

```

148 GCGGCAACAGCAGCAGTGTGGCGCAAAATCTGGAGGGGAGCCCGCAGGGTCCGCGCT 207
197 CGACCTGCTGTTGAAGCACAGCAGTGTGAGGGGCGCTCTGCTGCGGCGCAGGAGA 256
208 CGACCTGCTAGTGA-----AGCCAGTCAAGGGCGGCTCGACTGCGGCGAGS--AAA 259
257 TCACCCGACCAAGGAGGAGGCGCTTGAGTGTATCAACGGCTACATCCAGAAGATCAAGT 316
260 TCACCCGACCAAGGAGGAGGCGCTTGAGTGTATCAACGGCTACATCCAGAAGATCAAG 319
317 CGGAGAGAGGAGCTTTGAGTGTCTGCGCTCACAGTTCAGGACTGAGCTCAGCCAGG 376
320 CAGGAGAGAGGAGCTTTGAGTGTCTGCGCTCACAAATTCAGGACTGAGCTTCGCCAAG 379
377 CCAGGAGAGAGCTGGTGGCTTTCAGCAGAGTGTAGATGAGAGCCATTTCAAGACGCT 436
380 CCAGGAGAGAGCTGGTGGCTTTCAGCAGAGTGTAGATGAGAGCCATTTGAGAGCCCT 439
437 CGTTTGGCTCGGACCGGGGAGATGAGCGGGCGCTGTTCACGGATTCCGGCATTCACA 496
440 GGTTCGCGCGGAGCGGGGAGATGAGCGGAGATGAGCGGAGTTCACGGATTCCGGCATTC 499
497 TCATCTCCGCACTGAGTGGAGGAGCGGAGCGGCTGCGCTCGGGCGAGGCGAGGCG 556
500 TCAATGTCCGACCGGAGTGGAGTGGGGGCGGCGAGGCTGCGCTCGGG----- 548
557 GGCTAGGCGGCGGAGCTCCCTCGGCGCGCAGCGGAGTGGCGGAGAACCCCGCATCCCTGC 616
549 -----TTCCCGCTGCTGCTGGCCAGTGGCCGAGACCCCGCATCCCTGC 594
617 CACCGTCAACAGTATTTATTTCCCAAAATGCTGGTGGAGGGGCGCTTCCAGATTGGG 676
595 CACTGTCAACAGTATTTATTTTACCAAAATGCTGGGAGGGGCGCTTCCAGACTGGG 654
677 GGCCT-----GGGTCCCACTCCCTGTCATCCCGCTGCTGGGCT 718
655 GGCCTTTCAGACTGGGGCGCTGGGCTTCCCACTTCCCTGTCAGTCCCGCTGGGCT 714
719 GCGACCGCCAGATTTCCCTTAAGGAATGACTTCAGCAGGGGTGGAGGCTCCCAAGACC 778
715 GCGACCTCAAGACTTCCCAATTAAGAGTGTACTTCAGCATGGGAGGGCTCGCAGACC 774
779 CAGGCGAGTGTGGTGGAGGGGTGTCCAAAGAGAGGCTGTGTGAGCAGAGCGGCGCCG 838
775 CAGGCGAATGTGTGGAGGAGTGTTCAAAGAGAGAGTCTGTGTGAGCAGAGCGGCGCCCA 834
839 TGTCCCGCCAGGTGTGGAGGAGGAGTGTGAGGCGC--GAATTGTTTCTAGTTAGGCGACG 896
835 TGTCCCGCCAGGTGTGGAGGAGGAGTGTGAGGCGC--GAATTGTTTCTAGTTAGGCGACG 894
897 CTCTCTGTTCAGTGTGAGGAGGAGTGTGAGGCGC--GAATTGTTTCTAGTTAGGCGACG 955
895 CTCTCTGTTCAGTGTGAGGAGGAGTGTGAGGCGC--GAATTGTTTCTAGTTAGGCGACG 954
956 CTGTGAG--ACCTTTCACCCCGCAATTAAACCCAGAACCA 994
955 CTTTTCAGTGTGAGGAGGAGTGTGAGGCGC--GAATTGTTTCTAGTTAGGCGACG 994

```

```

RESULT 10
AL391728
LOCUS Human DNA sequence from clone RP11-423A18 on chromosome 1, complete
DEFINITION sequence.
ACCESSION AL391728
VERSION AL391728.19 GI:21217762
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 153750)

```

AUTHORS Wallis, J.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT On May 27, 2002 this sequence version replaced gi.21213126. During difference assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep. This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1>.
 RP11-423A18 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
 VECTOR: pBACE3.6. Location/Qualifiers
 1. .153750
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="11"
 /clone="RP11-423A18"
 /clone_lib="RP11-11.2"

FEATURES
 source
 Query Match 69.0%; Score 685.6; DB 9; Length 153750;
 Best Local Similarity 86.18; Pred. No. 1.1e-134;
 Matches 861; Conservative 0; Mismatches 84; Indels 55; Gaps 7;

ORIGIN

17 GAGGGAAGATGGCGGACGAGGAGAGTGGCGCGCGTGGGAGAACGCGATGAGCGCA 76
 Db 107991 GAGGGAAGATGGCGGACGAGGAGAACTGGCGCGTGGTGGGAAAGCGCATGAGCGGCC 108050

77 GCTCAGGCGAGTGACTACTTCAACCATCACTAAGCCGAGCGAGTGGGAGCGGCCA 136
 Db 108051 CCTCAGGCGAGGATGCTACTTCAACCATCACTAAGCCGAGCGAGTGGGAGCGGCCA 108110

137 GCGGCAACAGCAGCAGTGGTGGGCAAAACGGCGAGGGAGCGCTGCCAGGGTCCGCTGCT 196
 Db 108111 GCGGCAACAGCAGCAGTGGTGGGCAAAACGGCGAGGGAGCGCTGCCAGGGTCCGCTGCT 108170

197 GCGACCTGCTGGTGAAGCAGCAGCAGTCAAGCGCGCCCTCGTCTTGGCGGAGGAGAGA 256
 Db 108171 GCGACCTGCTAGTGA-----AGCCAGTCAAGCGCGCCCTCGACCTGGCGGAGG---AAA 108222

257 TCACCCGACCAAGAGAGGAGCGCCCTGGAGCTGATCAACGGCTACATCCAGAAAGATCAAGT 316
 Db 108223 TCACCCGACCAAGAGAGGAGCGCCCTGGAGCTGATCAGCGGCTACATCCAGAAAGATCAAGG 108282

317 CGGAGAGAGGAGCTTTGAGTCTTGGGCTCACAGTTCAGCGACTGCGAGCTCAGCCAAAG 376
 Db 108283 CAGGAGAGAGGAGCTTTGAGTCTTGGGCTCACAGTTCAGCGACTGCGAGCTCGGCCAAGG 108342

377 CCAGGGGAGACTGGGTGCTTTCAGCAGAGGTCAGATGAGGAGCCATTTGAAGACGGCT 436
 Db 108343 CCAGGGGAGACTGGGTGCTTTCAGCAGAGGTCAGATGAGGAGCCATTTGAAGACGGCT 108402

437 CGTTTGGCTCGGACCGGGGAGATGAGCGGGCCCGTGTTCACGATTCGGCATTCACAA 496
 Db 108403 GGTTCGGCGGCGAGCGGGGAGATGAGCGGGAGATGAGCGGATTCGGCATTCACAA 108462

497 TCATCTCCGACATGAGTGGGAGAGCGGAGCGCTGAGCTCGGGGAGGAGGCGGCG 556
 Db 108463 TCATCTCCGACATGAGTGGGAGAGCGGAGCGCTGAGCTCGGGGAGGAGGCGGCG 108511

557 GGCTAGGCGGCGGAGCTCCCTTCGCGCGAGCGAGTGGCGAAACCCCGACCTCCCTGC 616
 Db 108512 -----TTCCCGCGCTGCTGTCGCCAGTGGCGAAACCCCGACCTCCCTGC 108557

617 CACCGTCACAGATTTATTTCCCAATGGTGGAGGGGCGCTTCAGATTGGG 676
 Db 108558 CACTGTCACAGATTTATTTACCAAAATGGTGGAGGGGCGCTTCAGACTGGG 108617

677 GGCCCT-----GGGTCCCACTCCCTGTCATCCCGAGTTGGGCT 718
 Db 108618 GGCCCTTCAGACTGGGGCGCTGGGGTTCCCACTCCCTGTCATCCCGAGTTGGGCT 108677

719 GCGACGCGCAGATTCTCCCTTAAGGAATGACTTCAGCAGGGGTGGGAGGCTCCCGAG 778
 Db 108678 GCGACTTCAGACTTCCTTAAGGAGTTGACTTCAGCATGGGAGGAGGCTCCGAGACC 108737

779 CAGGGCAGTGTGGTGGGAGGGTGTTCAAAGAGAGGCGCTGTCAGCAGAGCGCGCCG 838
 Db 108738 CAGGGCAATGTGGTGGGAGGAGTGTTCAAAGAGAGAGTCTGTCAGCAGAGCGCGCCCA 108797

839 TGTCCCGCAGGTGCTGGAGGAGGAGTCCAGGCGCC--GAAATTGTTCTAGTTAGGCGCAG 896
 Db 108798 TGTCCCGCAGGTCTCTGGAGGAGAGTCCAGGCGCCCGAGGCTGTTTGTAGTTAGGCGCAG 108857

897 CTCTCTGTTCTAGTCGCAAGGTGAACACTCATGCG--GCAGCCATGGGCGCTCTGAGCAA 955
 Db 108858 CTCTCTGTTCTAGTACAAAGAACAAACACTCTGTCGGCCAGCTGTGGGCGCTCTGAGCAA 108917

956 CTGTGCAG--ACCTTTTACCCCCCAATTAACCCAGAACCA 994
 Db 108918 CTTTGCAGCACCTTTTACCCCCCAATTAATCGGAACCA 108957

RESULT 11
 AX834874
 LOCUS 1656 bp DNA linear PAT 15-DEC-2003
 DEFINITION Sequence 1998 from Patent EP1347046.
 ACCESSION AX834874
 VERSION AX834874.1 GI:39921009
 KEYWORDS Homo sapiens (human)
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 Isogai, T., Sugiyama, T., Otsuki, T., Wakamatsu, A., Sato, H., Ishii, S.,
 Yamamoto, J. I., Isono, Y., Hio, Y., Otsuka, K., Nagai, K., Irie, R.,
 Tamechika, I., Seki, N., Yoshikawa, T., Otsuka, M., Nagahari, K. and
 Masuho, Y.
 Full-length cDNA sequences
 Patent: EP 1347046-A 1998 24-SEP-2003;
 Research Association for Biotechnology (JP)
 Location/Qualifiers
 source
 1. .1656
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

FEATURES
 source
 1. .1656
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN

Query Match 57.1%; Score 567.4; DB 6; Length 1656;
 Best Local Similarity 98.5%; Pred. No. 1.6e-109;
 Matches 594; Conservative 0; Mismatches 6; Indels 3; Gaps 2;

395 CCTTCAGCAGAGGTGATGATGAGGAGCCATTTGAAGACGGCTCGTTGCGTGGGAGCG 454

Db	1045	CCTGGCTCCAGGTCAGATGCAGAGCCATTGTAAGACGGCTCGTTTGGCTGGGACGG	1104
Qy	455	GGGAGATGAGCGGCGCGTGTTCACGGATTCCGGCATCCACATCATCTCCGCACTGAGT	514
Db	1105	GGGAGATGAGCGGCGCGTGTTCACGGATTCCGGCATCCACATCATCTCCGCACTGAGT	1164
Qy	515	GAGGGTGGGAGCCAGCCTGGCTCGGGCGAGGCGAGGCGGCGTGGCGGCGGACGTC	574
Db	1165	GAGGGTGGGAGCCAGCCTGGCTCGGGCGAGGCGAGGCGGCGTGGCGGCGGACGTC	1224
Qy	575	CCCCTTCCCGCAGCAGTGGCGGAACCCCGCACTCCCTCCACCGTCACACAGTATT	634
Db	1225	CCCCTTCCCGCAGCAGTGGCGGAACCCCGCACTCCCTCCACCGTCACACAGTATT	1284
Qy	635	ATTGTTCCCAATGGCTGGGAGGGGCGCTTCCAGATTGGGGCGCTGGGGTCCCCACT	694
Db	1285	ATTGTTCCCAATGGCTGGGAGGGGCGCTTCCAGATTGGGGCGCTGGGGTCCCCACT	1344
Qy	695	CCCTGTCCATCCCGAGTGGGCTCGACCGCGAGATTCTCCCTTAAGCAATTGACTTCA	754
Db	1345	CCCTGTCCATCCCGAGTGGGCTCGACCGCGAGATTCTCCCTTAAGCAATTGACTTCA	1404
Qy	755	GCAGGGTGGGAGGCTCCCGAGACCCAGGCGAGTGTGGGAGGGCTGTTCACAAAGAA	814
Db	1405	GCAGGGTGGGAGGCTCCCGAGACCCAGGCGAGTGTGGGAGGGCTGTTCACAAAGAA	1464
Qy	815	GGCTGTGTCAGCAGACCGCCCGTGTCCCGCCAGGTCGTGGAGCGACATCGAGGCGG	874
Db	1465	GGCTGTGTCAGCAGACCGCCCGTGTCCCGCCAGGTCGTGGAGCGACATCGAGGCGG	1524
Qy	875	AATTGTTTCTAGTTAGGCGACCTCTCTGTTCAGTCCGCAAGGTGAACACTCATCGGC	933
Db	1525	AATTGTTTCTAGTTAGGCGACCTCTCTGTTCAGTCCGCAAGGTGAACACTCATCGGC	1584
Qy	934	-CAGCCATGGGCGCTCTGAGCAACTGTGAGCACCCTTCACCCCCAATTAAACCCAGAA	991
Db	1585	CCAGCCATGGGCGCTCTGAGCAACTGTGAGCACCCTTCACCCCCAATTAAACCCAGAA	1644
Qy	992	CCA 994	
Db	1645	CCA 1647	
RESULT 12			
AK097558			
LOCUS			
DEFINITION			
Homo sapiens cDNA FLJ40239 fis, clone TESTI203436, highly similar to PEPTIDYL-PROLYL CIS-TRANS ISOMERASE NIMA-INTERACTING 1 (EC 5.2.1.8).			
ACCESSION			
AK097558			
VERSION			
AK097558.1 GI:21757365			
KEYWORDS			
oligo capping; fis (full insert sequence).			
SOURCE			
Homo sapiens (human)			
ORGANISM			
Homo sapiens			
REFERENCE			
AUTHORS			
Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakanatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatesuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.			
TITLE			
NEDO human cDNA sequencing project			
JOURNAL			
Unpublished			
REFERENCE			
2 (bases 1 to 1656)			
AUTHORS			
Isogai,T. and Yamamoto,J.			
TITLE			
Direct Submission			
JOURNAL			
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)			

COMMENT		NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.			
FEATURES		Location/Qualifiers			
source		1. 1656			
		/organism="Homo sapiens"			
		/mol_type="mRNA"			
		/db_xref="taxon:9606"			
		/clone="TESTI203436"			
		/tissue_type="testis"			
		/clone_lib="TESTI2"			
		/note="cloning vector: pME18SFL3"			
ORIGIN					
Query Match		57.1%; Score 567.4; DB 9; Length 1656;			
Best Local Similarity		98.5%; Pred. No. 1.6e-109;			
Matches		594; Conservative 0; Mismatches 6; Indels 3; Gaps 2;			
Qy	395	CCTTCAGCAGAGTTCAGATGCAGAGCCATTGTAAGAGCGCTCGTTTCGCTGGGACGG	454		
Db	1045	CCTGGCTCCAGGTCAGATGCAGAGCCATTGTAAGAGCGCTCGTTTCGCTGGGACGG	1104		
Qy	455	GGGAGATGAGCGGCGCGTGTTCACGGATTCCGGCATCCACATCATCTCCGCACTGAGT	514		
Db	1105	GGGAGATGAGCGGCGCGTGTTCACGGATTCCGGCATCCACATCATCTCCGCACTGAGT	1164		
Qy	515	GAGGGTGGGAGCCAGCCTGGCTCGGGCGAGGCGAGGCGGCTAGGCGGCGGACGTC	574		
Db	1165	GAGGGTGGGAGCCAGCCTGGCTCGGGCGAGGCGAGGCGGCTAGGCGGCGGACGTC	1224		
Qy	575	CCCCTTCCCGCAGCAGTGGCGGAACCCCGCACTCCCTCCACCGTCACACAGTATT	634		
Db	1225	CCCCTTCCCGCAGCAGTGGCGGAACCCCGCACTCCCTCCACCGTCACACAGTATT	1284		
Qy	635	ATTGTTCCCAATGGCTGGGAGGGGCGCTTCCAGATTGGGGCGCTGGGGTCCCCACT	694		
Db	1285	ATTGTTCCCAATGGCTGGGAGGGGCGCTTCCAGATTGGGGCGCTGGGGTCCCCACT	1344		
Qy	695	CCCTGTCCATCCCGAGTGGGCTCGACCGCGAGATTCTCCCTTAAGCAATTGACTTCA	754		
Db	1345	CCCTGTCCATCCCGAGTGGGCTCGACCGCGCAGATTCTCCCTTAAGCAATTGACTTCA	1404		
Qy	755	GCAGGGTGGGAGGCTCCCGAGACCCAGGCGAGTGTGGGAGGGCTGTTCACAAAGAA	814		
Db	1405	GCAGGGTGGGAGGCTCCCGAGACCCAGGCGAGTGTGGGAGGGCTGTTCACAAAGAA	1464		
Qy	815	GGCTGTGTCAGCAGACCGCCCGTGTCCCGCCAGGTCGTGGAGCGACATCGAGGCGCG	874		
Db	1465	GGCTGTGTCAGCAGACCGCCCGTGTCCCGCCAGGTCGTGGAGCGACATCGAGGCGCG	1524		
Qy	875	AATTGTTTCTAGTTAGGCGCACCTCTCTGTTCAGTCCGCAAGGTGAACACTCATGCGG	933		
Db	1525	AATTGTTTCTAGTTAGGCGCACCTCTCTGTTCAGTCCGCAAGGTGAACACTCATGCGG	1584		
Qy	934	-CAGCCATGGGCGCTCTGAGCAACTGTGAGCACCCTTCACCCCCAATTAAACCCAGAA	991		
Db	1585	CCAGCCATGGGCGCTCTGAGCAACTGTGAGCACCCTTCACCCCCAATTAAACCCAGAA	1644		
Qy	992	CCA 994			
Db	1645	CCA 1647			
RESULT 13		138036 bp DNA linear PRI 13-JUL-2002			
AC008752/c		AC008752			
LOCUS		Homo sapiens chromosome 19 clone CTD-2623N2, complete sequence.			
DEFINITION		AC008752			
ACCESSION		AC008752			

Tue Jun 29 08:37:21 2004

u49070.rge

Page 13

Search completed: June 28, 2004, 16:23:11
Job time : 4369.63 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2004, 14:35:24 ; Search time 545.075 Seconds
(without alignments)
7747.025 Million cell updates/sec

Title: U49070

Perfect score: 994

Sequence: 1 TCTGTGCCAGCACCCTCGAGG.....CCCAATTAAACCCAGAACCA 994

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Match	Length	DB ID	Description
1	994	100.0	994	5	AAH76459	Aah76459 cDNA corr
2	994	100.0	1014	2	AAT68888	Aat68888 NIVA-inte
3	892.6	89.8	3125	9	ADB63201	ADB63201 Human CDN
4	892.6	89.8	3143	4	AAD05229	Aad05229 Human sec
5	887.8	89.3	1143	6	ABK35196	ABK35196 Human CDN
6	729	73.3	1034	5	AA576759	AA576759 DNA encod
7	593.8	59.7	1045	5	AA576134	AA576134 DNA encod
8	492.2	49.5	580	2	AZ33510	Az33510 Human pro
9	392	39.4	458	8	ACH32387	Ach32387 Human end
10	217.6	21.9	559	4	AA62612	AA62612 Human foe
11	217.6	21.9	559	4	AA142606	AA142606 Probe #11
12	217.6	21.9	559	4	AAK36818	AAK36818 Human bon
13	217.6	21.9	559	4	AAK10973	AAK10973 Human bra
14	217.6	21.9	559	6	ABS10823	ABS10823 Human gen
15	213	21.4	213	4	ABA75131	ABA75131 Human foe
16	213	21.4	213	4	AA15672	AA15672 Probe #24
17	213	21.4	213	4	AAK49765	AAK49765 Human bon
18	213	21.4	213	4	AAK23649	AAK23649 Human bra
19	213	21.4	213	6	ABS23259	ABS23259 Human gen
20	208.4	21.0	439	6	AB199260	AB199260 Mouse isc
21	203.2	20.4	314	4	AAf92335	AAf92335 Bovine ma
22	200.8	20.2	296	2	AAQ60623	AAQ60623 Human bra
23	178	17.9	601	9	ADB55060	ADB55060 Toxicity-

C	24	178	17.9	601	9	ADB49564	ADB49564 Primary r
	25	161.4	16.2	995	4	ABL18667	ABl18667 Drosophil
	26	112	11.3	468	5	AA576133	AA576133 DNA encod
	27	110.4	11.1	537	7	ABT19273	ABt19273 Aspergill
	28	110	11.1	345	5	AA576758	AA576758 DNA encod
	29	108.2	10.9	406	3	AAf11336	AAf11336 Aspergill
	30	107.2	10.8	537	7	ABT21093	ABt21093 Aspergill
C	31	98.4	9.9	3475	4	ABL18666	ABl18666 Drosophil
	32	98.4	9.9	5330	4	ABL03148	ABl03148 Drosophil
C	33	98.4	9.9	7208	4	ABL02698	ABl02698 Drosophil
	34	84.6	8.5	366	7	AB253193	AB253193 Aspergill
	35	84.4	8.5	285	6	ABN79537	ABn79537 Human iso
	36	80	8.0	833	3	AAc47563	AAc47563 Arabidops
	37	78.4	7.9	747	3	AAc33551	AAc33551 Arabidops
	38	78.4	7.9	835	3	AAc36658	AAc36658 Arabidops
	39	74.8	7.5	662	3	AAf07816	AAf07816 Fusarium
	40	67.6	6.8	630	9	ADB69628	ADB69628 C. nesfor
	41	66	6.6	652	7	ABT18679	ABt18679 Aspergill
	42	66	6.6	652	7	ABT20495	ABt20495 Aspergill
	43	66	6.6	2652	7	ABT19899	ABt19899 Aspergill
	44	66	6.6	2652	7	ABT18085	ABt18085 Aspergill
C	45	66	6.6	5360	4	ABL20104	ABl20104 Drosophil

ALIGNMENTS

RESULT 1
AAH76459
ID AAH76459 standard; cDNA; 994 BP.
XX
AC AAH76459;
DT 22-OCT-2001 (first entry)
XX
DE cDNA corresponding to human IFN-alpha induced gene encoding PIN-1.

XX Human; interferon-alpha induced gene; type I interferon treatment;
KW chronic viral hepatitis; relapsing remitting multiple sclerosis;
KW neoplastic disease; IFN-alpha; interferon-alpha; PIN-1; ss.
XX

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 25..516

FT /*tag= a

FT /product= "PIN-1"

XX WO200159155-A2.

XX 16-AUG-2001.

XX 09-FEB-2001; 2001WO-GB000578.

XX 11-FEB-2000; 2000GB-00003203.

XX 11-FEB-2000; 2000GB-00003204.

XX 11-FEB-2000; 2000GB-00003205.

XX 11-FEB-2000; 2000GB-00003206.

XX 11-FEB-2000; 2000GB-00003207.

XX 11-FEB-2000; 2000GB-00003208.

XX 11-FEB-2000; 2000GB-00003210.

XX 11-FEB-2000; 2000GB-00003212.

XX 11-FEB-2000; 2000GB-00003213.

XX 11-FEB-2000; 2000GB-00003215.

XX 11-FEB-2000; 2000GB-00003216.

XX 11-FEB-2000; 2000GB-00003219.

XX 11-FEB-2000; 2000GB-00003220.

XX 11-FEB-2000; 2000GB-00003221.

XX 11-FEB-2000; 2000GB-00003222.

XX 17-FEB-2000; 2000GB-00003768.

XX (PHAR-) PHARMA PACIFIC PTY LTD.

XX

CC recombinant expression vector comprising the DNA sequence and host cells
 CC containing the vector are claimed. Methods are also claimed for
 CC identifying proteins that inhibit the mitosis promoting function of NIMA
 CC protein kinase and for controlling the growth of a cell by reducing Pini
 CC activity or Pini expression using an inhibitor, anti-Pini antibody,
 CC antisense nucleotide sequence or ribozyme, or by increasing Pini activity
 CC in the presence of an activator or increasing Pini expression using an
 CC enhancer. This allows treatment of cell proliferation disorders such
 CC adenocarcinomas, cancers, psoriasis, pemphigus vulgaris, acute
 CC respiratory distress syndrome, rheumatoid arthritis, septic shock and
 CC inflammation
 XX

SQ Sequence 1014 BP; 220 A; 310 C; 324 G; 160 T; 0 U; 0 Other;

Query Match 100.0%; Score 994; DB 2; Length 1014;
 Best Local Similarity 100.0%; Pred. No. 7.9e-230;
 Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCTGGCCAGCAGCTCGAGGGAAGATGCGGACGAGGAGAGCTGCCCGCGCTGGGAG 60
 DB 1 TGCTGGCCAGCAGCTCGAGGGAAGATGCGGACGAGGAGAGCTGCCCGCGCTGGGAG 60
 QY 61 AAGCGCATGAGCCGAGCTCAGGCGAGTGACTACTTCAACCAATCACTAAGCCAGC 120
 DB 61 AAGCGCATGAGCCGAGCTCAGGCGAGTGACTACTTCAACCAATCACTAAGCCAGC 120
 QY 121 CAGTGGGAGCGGCCAGCGGCAACAGCAGCAGTGCTGGCAAAAGCGGCGGGAGCCT 180
 DB 121 CAGTGGGAGCGGCCAGCGGCAACAGCAGCAGTGCTGGCAAAAGCGGCGGGAGCCT 180
 QY 181 GCCAGGCTCGCTGCTCGCAGCTCTGCTGTAAGCAGCAGCAGTCAACGCGGCGCTCGTCC 240
 DB 181 GCCAGGCTCGCTGCTCGCAGCTCTGCTGTAAGCAGCAGCAGTCAACGCGGCGCTCGTCC 240
 QY 241 TGGCGGCGAGGAGAGATCACCCGGACCAAGGAGGAGGCCCTGGAGCTCATCAAGCGCTAC 300
 DB 241 TGGCGGCGAGGAGAGATCACCCGGACCAAGGAGGAGGCCCTGGAGCTCATCAAGCGCTAC 300
 QY 301 ATCCAGAGATCAAGTCGGGAGAGAGAGACTTTGAGTCTCTGGGCTCAGCTGAGGAC 360
 DB 301 ATCCAGAGATCAAGTCGGGAGAGAGAGACTTTGAGTCTCTGGGCTCAGCTGAGGAC 360
 QY 361 TGCAGCTCAGCAAGCCAGCGGAGCACTGGTGCTTTCAGCAGAGGTGAGATGAGGAG 420
 DB 361 TGCAGCTCAGCAAGCCAGCGGAGCACTGGTGCTTTCAGCAGAGGTGAGATGAGGAG 420
 QY 421 CCATTGAGAGCGCTCGTTTGGCTGCGGACGCGGAGAGATGAGCGGCGCGTGTTCAG 480
 DB 421 CCATTGAGAGCGCTCGTTTGGCTGCGGACGCGGAGAGATGAGCGGCGCGTGTTCAG 480
 QY 481 GATTCCGGCATCCATCATCTCCGCACTGAGTGAGGTGGGAGGCCAGGCGCTGGCT 540
 DB 481 GATTCCGGCATCCATCATCTCCGCACTGAGTGAGGTGGGAGGCCAGGCGCTGGCT 540
 QY 541 CGGGCAGGCGAGGCGGCTAGCGCGGCGAGCTCCCTTCCCGCCAGCCAGTGGCCGA 600
 DB 541 CGGGCAGGCGAGGCGGCTAGCGCGGCGAGCTCCCTTCCCGCCAGCCAGTGGCCGA 600
 QY 601 ACCCGCCACTCCCTGCGCACCGTCAACAGTATTTATTTGTTCCCAATGGCTGGAGGG 660
 DB 601 ACCCGCCACTCCCTGCGCACCGTCAACAGTATTTATTTGTTCCCAATGGCTGGAGGG 660
 QY 661 GCCCTTCCAGATTGGGGGCGCTGGGCTCCCACTCCCTGCTCCATCCCAAGTTGGGGCTGC 720
 DB 661 GCCCTTCCAGATTGGGGGCGCTGGGCTCCCACTCCCTGCTCCATCCCAAGTTGGGGCTGC 720
 QY 721 GACCGCCAGATTCTCCCTTAAAGAAATTGACTTTCAGCAGGGGTGGAGGCTCCAGACCCA 780
 DB 721 GACCGCCAGATTCTCCCTTAAAGAAATTGACTTTCAGCAGGGGTGGAGGCTCCAGACCCA 780
 QY 781 GGGCAGTGTGGAGGGGTGTTCAAGAGAGAGGCTGGTTCAGCAGAGCGCCCGCTGG 840
 DB 781 GGGCAGTGTGGAGGGGTGTTCAAGAGAGAGGCTGGTTCAGCAGAGCGCCCGCTGG 840

QY 841 TCCCCCAGGTGCTGGAGGAGAGTCTGAGGGCCGAATTGTTCTAGTTAGGCCACGCTCC 900
 DB 841 TCCCCCAGGTGCTGGAGGAGAGTCTGAGGGCCGAATTGTTCTAGTTAGGCCACGCTCC 900
 QY 901 TCTGTTTCTGAGTTCGCAAGGTGAACACTCATGCGGAGCCATGCGGCGCTCTGAGCAACTGTG 960
 DB 901 TCTGTTTCTGAGTTCGCAAGGTGAACACTCATGCGGAGCCATGCGGCGCTCTGAGCAACTGTG 960
 QY 961 CAGACCTTTTACCCCCCAATTAACCCAGAACCA 994
 DB 961 CAGACCTTTTACCCCCCAATTAACCCAGAACCA 994

RESULT 3

ADB63201

ID ADB63201 standard; cDNA; 3125 BP.

XX ADB63201;

XX AC ADB63201;

XX DT 04-DEC-2003 (first entry)

XX DE Human cDNA encoding clone SPLEN20136700.

XX KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
 KW tissue regeneration; cell regeneration; membrane protein;
 KW signal transduction-related protein; transcription-related protein;
 KW osteoporosis; neurological disease; cancer; tumour.

XX OS Homo sapiens.

XX XX Key Location/Qualifiers
 CDS 1296..1628
 /*tag= a
 /product= "Clone SPLEN20136700 protein"

XX FT

XX PN EP1308459-A2.

XX PD 07-MAY-2003.

XX PF 28-MAR-2002; 2002EP-00007401.

XX PR 05-NOV-2001; 2001JP-00379298.

XX PR 25-JAN-2002; 2002US-00350978.

XX XX (HELI-) HELIX RES INST.

XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.

XX PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 XX WPI; 2003-450961/43.
 DR P-PSDB; ADB65171.

XX PT New polynucleotides and polypeptides, useful for developing a diagnostic
 PT marker or medicines for regulation of their expression and activity, or
 PT as targets of gene therapy.
 XX PS Claim 1; Page; 222pp; English.

XX CC The invention discloses a polynucleotide comprising a sequence selected
 CC from 1970 fully defined nucleotide sequences which encode novel
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 CC or its partial peptide, an antibody binding to the polypeptide or peptide
 CC of the polynucleotide, immunologically assaying the polypeptide or
 CC peptide of the polynucleotide by contacting the polypeptide or peptide
 CC with the antibody of the encoded protein, and observing the binding
 CC between the two, a transformant carrying the polynucleotide in an
 CC expressible manner and an antisense polynucleotide. The oligonucleotide
 CC is useful as a primer for synthesising the polynucleotide, or as a probe
 CC for detecting the polynucleotide. The polynucleotides and encoded
 CC proteins are useful as pharmaceutical agents and many disease-related

CC genes may be included in them, for developing a diagnostic marker or
 CC medicines for regulation of their expression and activity, or as targets
 CC of gene therapy. The genes are involved in tissue and/or cell
 CC regeneration. Membrane proteins, signal transduction-related proteins,
 CC transcription-related proteins, disease-related proteins and genes
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
 CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
 CC the activity or expression of the encoded protein to treat diseases. The
 CC sequence presented is a cDNA of the invention. Note: Some of the sequence
 CC data for this patent is not represented in the printed specification, but
 CC is based on sequence information supplied by the European Patent Office.

XX Sequence 3125 BP; 622 A; 893 C; 805 G; 805 T; 0 U; 0 Other;

Query Match 89.8%; Score 892.6; DB 9; Length 3125;
 Best Local Similarity 98.7%; Pred. No. 2.9e-205;
 Matches 921; Conservative 0; Mismatches 9; Indels 3; Gaps 2;

QY	65	GCATGAGCCGACGCTCAGCGCGAGTGTACTTCAACACATCACTAACGCCAGCCAGT	124
DB	2191	GCCTGCCCTCCCTCCAGCGCGAGTGTACTTCAACACATCACTAACGCCAGCCAGT	2250
QY	125	GGGAGCGGCCGACGCGCAACAGCAGCAGTGTGGCGCAAAACGGCGAGGGGAGCCCTGCCA	184
DB	2251	GGGAGCGGCCGACGCGCAACAGCAGCAGTGTGGCGCAAAACGGCGAGGGGAGCCCTGCCA	2310
QY	185	GGTCCGCTGTCTGCACTCTCTCTGTAAGCAGCAGTCAAGCGCGCCCTCTCTGGC	244
DB	2311	GGTCCGCTGTCTGCACTCTCTGTAAGCAGCAGTCAAGCGCGCCCTCTCTGGC	2370
QY	245	GGCAGGAGAGATCACCCGCGACCAAGAGGAGGCCCTGGAGCTGATCAACGGGTACATCC	304
DB	2371	GGCAGGAGAGATCACCCGCGACCAAGAGGAGGCCCTGGAGCTGATCAACGGGTACATCC	2430
QY	305	AGAAGATCAAGTCGGGAGAGAGACTTTTCAGTCTCTGGCTCAGATTCAGCGACTGCA	364
DB	2431	AGAAGATCAAGTCGGGAGAGAGACTTTTCAGTCTCTGGCTCAGATTCAGCGACTGCA	2490
QY	365	GCTCAGCAAGGCGAGGGAGACTGGTGCCTTCAGCAGAGTCAGATCAGAGGCCAT	424
DB	2491	GCTCAGCAAGGCGAGGGAGACTGGTGCCTTCAGCAGAGTCAGATCAGAGGCCAT	2550
QY	425	TTGAGACGCTCTGTTTGGCTGGGAGGAGATGAGCGGGCCCGTGTTCACGGATT	484
DB	2551	TTGAGACGCTCTGTTTGGCTGGGAGGAGATGAGCGGGCCCGTGTTCACGGATT	2610
QY	485	CCGGCATCCATCATCTCCGCACTGAGTGGGAGGAGCCAGGCTGGCTGGGCTGGG	544
DB	2611	CCGGCATCCATCATCTCCGCACTGAGTGGGAGGAGCCAGGCTGGCTGGGCTGGG	2670
QY	545	GCAGGCGAGGCGGCTAGCGCGGACGCTCCCTTCCCGCCAGCGAGTGGCGGACCC	604
DB	2671	GCAGGCGAGGCGGCTAGCGCGGACGCTCCCTTCCCGCCAGCGAGTGGCGGACCC	2730
QY	605	CCCACTCCCTGCCACCGCTCAGCAGTATTTATTTCTCCCAATGGCTGGAGGGGCGCC	664
DB	2731	CCCACTCCCTGCCACCGCTCAGCAGTATTTATTTCTCCCAATGGCTGGAGGGGCGCC	2790
QY	665	TTCCAGATTGGGGGCGCTGGGGTCCCACTCCCTGTCTCATCCAGTTGGGCTGGCACC	724
DB	2791	TTCCAGATTGGGGGCGCTGGGGTCCCACTCCCTGTCTCATCCAGTTGGGCTGGCACC	2850
QY	725	GCCAGATTCTCCCTTAAGGAATTCAGTTCAGCAGGGGTGGAGGCTCCAGACCCAGGGC	784
DB	2851	GCCAGATTCTCCCTTAAGGAATTCAGTTCAGCAGGGGTGGAGGCTCCAGACCCAGGGC	2910
QY	785	AGTGTGGTGGAGGGGTGTTCCAAAGAGAGAGCCCTGTGTACAGCAGCGCCCGCTGTCCC	844
DB	2911	AGTGTGGTGGAGGGGTGTTCCAAAGAGAGAGCCCTGTGTACAGCAGCGCCCGCTGTCCC	2970
QY	845	CCGAGTCTGGAGGACAGCTCAGGGGCGGAATTTCTAGTTAGGACAGCTCTCTCTG	904
DB	2971	CCGAGTCTGGAGGACAGCTCAGGGGCGGAATTTCTAGTTAGGACAGCTCTCTCTG	3030

QY	905	TTCACTGCGAAAGGTGAACACTCATGGG--CAGCCATGGCCCTCTCAGCAACTGTGCA	962
DB	3031	TTCACTGCGAAAGGTGAACACTCATGGG--CAGCCATGGCCCTCTCAGCAACTGTGCA	3090
QY	963	G-ACCTTTTACCCCGCAATTAAACCCAGAACCA	994
DB	3091	GCACCTTTACCCCGCAATTAAACCCAGAACCA	3123

RESULT 4

AAD05229

ID AAD05229 standard; cDNA; 3143 Bp.

XX AAD05229;

AC AAD05229;

XX 18-JUL-2001 (first entry)

XX Human secreted protein-encoding gene 10 cDNA clone HHEPJ23, SEQ ID NO:20.

XX Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
 XX foetal abnormality; developmental abnormality; haematopoietic disorder;
 XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
 XX Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
 XX psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
 XX inflammation; neurological disorder; Alzheimer's disease; food additive;
 XX angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
 XX pregnancy-related disorder; endocrine disorder; infection; wound healing;
 XX cell culture; chemotaxis; vulnery; binding partner identification;
 XX gene therapy; chromosome 19; ss.

XX Homo sapiens.

XX Location/Qualifiers

XX CDS 175..363

XX /tag= a

XX /product= "Human secreted protein"

XX sig_peptide 175..237

XX /tag= b

XX mat_peptide 238..360

XX /tag= c

XX /product= "Mature human secreted protein"

XX WO200134629-A1.

XX 17-MAY-2001.

XX 08-NOV-2000; 2000WO-US030654.

XX 12-NOV-1999; 99US-0164835P.

XX 27-JUL-2000; 2000US-0221142P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Komatsoulis GA, Wei P, Fiscella M, Baker KP;

XX WPI; 2001-308779/32.

XX P-PSDB; AAE01361.

XX New nucleic acid encoding one of 21 human secreted proteins for
 XX diagnosing, preventing, treating or ameliorating medical conditions, such
 XX as autoimmune disease and cancer, and used as a food additive or
 XX preservative.

XX Claim 1; Page 386-387; 490pp; English.

XX AAD05220-AAD05282 represent cDNAs corresponding to 21 human secreted
 XX protein genes, and AAE01352-AAE01413 represent the proteins they encode.
 XX AAE01415-AAE01433 represent human secreted protein fragments or variants.
 XX The secreted proteins and their genes are useful for preventing, treating
 XX or ameliorating medical conditions, e.g., by protein or gene therapy.
 XX Pathological conditions can be diagnosed by determining the amount of the
 XX new protein in a sample or by determining the presence of mutations in

the new genes. Specific uses are described for each of the 21 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, haematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angiogenic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infectious cell proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein-encoding cDNA of the invention

Sequence 3143 BP; 678 A; 888 C; 802 G; 762 T; 0 U; 13 Other;

Query Match 89.8%; Score 892.6; DB 4; Length 3143;
Best Local Similarity 98.7%; Pred. No. 3e-205;
Matches 921; Conservative 0; Mismatches 9; Indels 3; Gaps 2;

QY 65 GCATGAGCCGAGCTCAGGCGAGTGACTTCAACCATCATCACTAAAGCCAGCCAGT 124
DB 2089 GCCTGCCCTCCCTCAGGCGAGTGACTTCAACCATCATCACTAAAGCCAGCCAGT 2148

QY 125 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGTGCAAAACGGCGAGGGAGCTGCCA 184
DB 2149 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGTGCAAAACGGCGAGGGAGCTGCCA 2208

QY 185 GGGTCCGCTGTCTGCACCTCTGTGTGAAGCAGCAGTCAAGCGGCGCCCTCCCTGGC 244
DB 2209 GGGTCCGCTGTCTGCACCTCTGTGTGAAGCAGCAGTCAAGCGGCGCCCTCCCTGGC 2268

QY 245 GGCAGGAGAGATCAACCGGACCAAGGAGAGGCGCTCGAGCTGATCAACGGCTACATCC 304
DB 2269 GGCAGGAGAGATCAACCGGACCAAGGAGAGGCGCTCGAGCTGATCAACGGCTACATCC 2328

QY 305 AGAAGATCAAGTCGGGAGAGAGGACTTTGAGTCTCTGGCCCTCAGAGTTCAGCGACTGCA 364
DB 2329 AGAAGATCAAGTCGGGAGAGAGGACTTTGAGTCTCTGGCCCTCAGAGTTCAGCGACTGCA 2388

QY 365 GCTCAGCCAGCGGAGGAGACCTGGGTGCTTCAGCAGAGGTCAGATGAGAGCCAT 424
DB 2389 GCTCAGCCAGCGGAGGAGACCTGGGTGCTTCAGCAGAGGTCAGATGAGAGCCAT 2448

QY 425 TTGAAGAGCGCTCGTTTTCGCTGCGGAGGAGAGTCAAGCGGCGCCGCTGTTTCAGCGATT 484
DB 2449 TTGAAGAGCGCTCGTTTTCGCTGCGGAGGAGAGTCAAGCGGCGCCGCTGTTTCAGCGATT 2508

QY 485 CCGGATCCATCATCATCTCCGACATGAGTGGGTGGGAGCCAGGCGCTGGCTCGG 544
DB 2509 CCGGATCCATCATCATCTCCGACATGAGTGGGTGGGAGCCAGGCGCTGGCTCGG 2568

QY 545 GCAGGCGAGGCGGCTAGCGCGGCGAGCTCCCTTGGCCCGCAGCGAGTGGCCGAACCC 604
DB 2569 GCAGGCGAGGCGGCTAGCGCGGCGAGCTCCCTTGGCCCGCAGCGAGTGGCCGAACCC 2628

QY 605 CCCACTCCCTCGCCACCGTCACACAGTATTTATTTGTTCCACAAATGCTGGGAGGGGCC 664
DB 2629 CCCACTCCCTCGCCACCGTCACACAGTATTTATTTGTTCCACAAATGCTGGGAGGGGCC 2688

QY 665 TTCAGATTTGGGGGCCCTCGGGTCCCACTCCCTGTCCATCCCACTGGGGTGGGAGCC 724
DB 2689 TTCAGATTTGGGGGCCCTCGGGTCCCACTCCCTGTCCATCCCACTGGGGTGGGAGCC 2748

QY 725 GCCAGATTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGGAGGCTCCAGACCCAGGGC 784
DB 2749 GCCAGATTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGGAGGCTCCAGACCCAGGGC 2808

QY 785 AGTGTGGTGGGAGGGGTGTTCAAGAGAGAGGCTGTGTCAGCAGAGCCGCCCGTGTCCC 844
DB 2809 AGTGTGGTGGGAGGGGTGTTCAAGAGAGAGGCTGTGTCAGCAGAGCCGCCCGTGTCCC 2868

QY 845 CCCAGTGTCTGAGGCGAGACTCGAGGCGCGAATTGTTCTAGTTAGGCCACGCTCCTCTG 904
DB 2869 CCCAGTGTCTGAGGCGAGACTCGAGGCGCGAATTGTTCTAGTTAGGCCACGCTCCTCTG 2928

QY 905 TTCAGTGCAGAAAGGTGAACACTCATCGCG--CAGCCATGGGCCCTCTGAGCAACTGTGCA 962
DB 2929 TTCAGTGCAGAAAGGTGAACACTCATCGCGCCAGCCAGCCATGGGCCCTCTGAGCAACTGTGCA 2988

QY 963 G-ACCTTTTACCCCCCAATTAACCCAGAACCA 994
DB 2989 GCACCTTTTACCCCCCAATTAACCCAGAACCA 3021

RESULT 5
ABK35196
ID ABK35196 standard; cDNA; 1143 BP.
XX
AC ABK35196;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human cDNA encoding secreted protein #334.
XX
KW Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
KW viral infection; bacterial infection; fungal infection; diabetes; asthma;
KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;
KW tissue regeneration; wound healing; burn; haematopoiesis;
KW myeloid cell deficiency; lymphoid cell deficiency.
OS Homo sapiens.
XX
FN WO200177288-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001WO-US010224.
XX
PR 06-APR-2000; 2000US-0195582P.
XX
(GEMY) GENETICS INST INC.
XX
PI Wong GS, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
WI WI; 2002-179321/23.
XX
PT Five hundred and ninety two polynucleotides derived from a variety of
PT human tissue sources which encode secreted proteins, useful for treating
PT immune deficiencies and disorders such as autoimmune disorders.
XX
PS Claim 1; Page 246; 372pp; English.
XX
CC The invention relates to 592 polynucleotides which have been derived from
CC a variety of human tissue sources and which encode novel secreted
CC proteins. The polynucleotides can be used as probes for the
CC identification and isolation of full length cDNA and genomic DNA. The
CC polynucleotides and proteins can also be used as nutritional supplements.
CC The proteins are useful in the treatment of various immune deficiencies
CC and disorders such as viral infections, bacterial infections, fungal
CC infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
CC sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
CC and conditions (e.g. asthma). They are also useful for treating

Db	86	AGTCCGAGTGTACTTCTTCAACCATCATCTAAACGACGACGAGTGGAGCGGCGCCATCGG	145
Qy	141	CAACAGCAGCAGTGTGTGGCAAAAAACGGCAGGGGAGCGCTGCCAGGTCGCTGCTGCTCGCA	200
Db	146	CAACAGCAGCAGTGTGTGGCAAAAAACGGCAGGGGAGCGCTGCCAGGTCGCTGCTGCTCGCA	205
Qy	201	CTGTCTGTGTGAAGCACAGCCAGTCAGCGGGGGCCCTCGTCTCTGGCGGCAGGAGAAGATCAC	260
Db	206	CTGTCTGTGTGAAGCACAGCCAGTCAGCGGGGGCCCTCGTCTCTGGCGGCAGGAGAAGATCAC	265
Qy	261	CGGACCAAGAGGAGGA-GGCCCTTGAGACTGATCAACGGCTACATCCAGAGATCAAGTCCG	319
Db	266	CGGACCAAGAGGAGGAGCGCCCTTGAGCTGATCAACGGCTACATCCAGAGATCAAGTCCG	325
Qy	320	GAGAGCAGACACTTTG-AGTCTCTGGCTCACAGTTAGCGACTGCGAGCTCAGCCAAAGGCC	378
Db	326	GAGAGCAGACACTTTGAGATCTCTGGCTCACAGTTAGCGACTGCGAGCTCAGCCAAAGGCC	385
Qy	379	-AGGGAGACCTGGG-TGCCCTTCAGCAGAGGTGAGATGCAGAGCCATTT-----GAAGA	431
Db	386	AAGGGAGACCTGGGTTCCTTCAGCAGAGGTGAGATGCAGAGCCATTTTGAGAACGC	445
Qy	432	CGCCTCGTTGCGCTGCGAGCGGGGAGATGAGCGGGCCCGTGTTCACGATTTCCGG	488
Db	446	CTCGTTTTCGGCTTGCGGACGGGGGAGATGAGCGGGCCCGTGTTCACGATTTCCGG	505
Qy	489	CATCCACATCATCCT--CCGCACATGATGAGGGTGGGGAGCCAGGCTTGCTTCGGGGC	546
Db	506	CATCCACATCATCCTCCCGACATTGATGAGGGTGGGAAGCCAGGCTTGCTTCGGGGC	565
Qy	547	AGGCGAGGGCGGCTAGGCCGGCGAGCTCCCGCTTCCCGCCAGCAGCTGCGCCGAA-CCCC	605
Db	566	AGGCGAGGGCGGCTAGGCCGGCGAGCTCCCGCTTCCCGCCAGCAGCTGCGCCGAAACCCCC	625
Qy	606	CCATCCCTCCGACCGTCACACATATTTATGTTCCACAATGGCTGGAGGGGGCGCT	665
Db	626	CCATCCCTCCGACCGTCACACATATTTATGTTCCACAATGGCTGGAGGGGGCGCT	685
Qy	666	TCCAGATTGGGGCCCTGGGTCGCCACTCCCTGTCATCCCAAGTTGGGGCTGCGACCG	725
Db	686	TCCAGATTGGGGCCCTGGGTCGCCACTCCCTGTCATCCCAAGTTGGGGCTGCGACCG	745
Qy	726	CCAGATTCTCCCTTAAAG-ATTTCACATTCACAGGGTGGAGAGCTCCACAGCCACAGGC	784
Db	746	CCAGATTCTCCCTTAAAGAAATTGACTTCAGCGGGTGGAGAGCTCCACAGCCACAGGC	805
Qy	785	AGTGTGTGGAGGGGTGTTCCAAAGAGAAGCCCTGGTCAGCAGAGCGCCCGGTGCC	844
Db	806	AGTGTGTGGAGGGGTGTTCCAAAGAGAAGCCCTGGTCAGCAGAGCGCCCGGTGCC	865
Qy	845	CCGAGGTGCTGGAGCAGACTCGAGGGCCGAATTGTTTCTAGTTAGGCCACGCTCCTCTG	904
Db	866	CCGAGGTGCTGGAGCAGACTCGAGGGCCGAATTGTTTCTAGTTAGGCCACGCTCCTCTG	925
Qy	905	TTCACTGCGCAAAAGGTGAACACTCATCGGG--CAGCCATGGGCCCTCTGAGCAACTGTGCA	962
Db	926	TTCACTGCGCAAAAGGTGAACACTCATCGGGCCAGCCATGGGCCCTTTGAGCACTGTGCA	985
Qy	963	G-ACCTTTTCACCCCAATTAAACCCAGAACCA	994
Db	986	GCACCTTTTCACCCCAATTAAACCCAGATCCA	1018

RESULT 7	
AAS76134	
ID	AAS76134 standard; cDNA; 1045 BP.
XX	
AC	AAS76134;
XX	
DT	13-FEB-2002 (first entry)
XX	
DE	DNA encoding novel human diagnostic protein #11938.

	QY	Db	QY	Db
199	CACCTGCTGGTGAAGCACACACGCCAGTCAACGCG-CGGCCTCGTCTCGTGGCGCAGGAGAAGAT	257		
207	CACCTGCTGGTGAAGCACACACGCCAGTCAACGCGGGCGCCTCGTCTCGTGGCGCAGGAGAAGAT	266		
258	CACCGGACCAAGGGAGGGCGCCCTGGAGCTGATCAACGGCTACATCCAGAAGATCAAGTC	317		
267	CACCGGACCAAGGGAGGGCGCCCTGGAGCTGATCAACGGCTACATCCAGAAGATCCAGTC	326		

ACH32387
ID ACH32387 standard; cDNA; 458 BP.
XX AC
XX ACH32387;
XX
DT 13-OCT-2003 (first entry)
XX
XX Human endothelial cell cDNA #520.
XX
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
XX genome mapping; biodiversity; genetic disorder.
XX
XX Homo sapiens.
XX
XX US2003073623-A1.
XX
XX 17-APR-2003.
XX
XX 30-JUL-2001; 2001US-00918995.
XX
XX 30-JUL-2001; 2001US-00918995.
XX
XX (DRMA/) DRMANAC R T.
XX (LABA/) LABAT I.
XX (STAC/) STACHE-CRAIN B.
XX (DICK/) DICKSON M C.
XX (JONE/) JONES L W.
XX
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
XX as hybridization probes, as oligomers for PCR, for chromosome and gene
XX mapping, in the recombinant production of protein, or in generating
XX antisense DNA or RNA.
XX
XX Claim 1; SEQ ID NO 19599; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
XX 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
XX determined by the technique of SBH (sequencing by hybridisation). Also
XX included is a purified polypeptide comprising a sequence corresponding to
XX a reading frame of the novel polynucleotide. The nucleic acid sequences
XX are useful in diagnostics as expressed sequence tags (EST) for
XX identifying expressed genes or for physical mapping of the human genome,
XX in forensics, in assessing biodiversity, or in identifying mutations
XX responsible for genetic disorders and other traits. The nucleotide
XX sequences are also useful as hybridisation probes, as oligomers for PCR,
XX for chromosome and gene mapping, in the recombinant production of
XX protein, or in generating antisense DNA or RNA. The purified polypeptide
XX is useful for generating antibodies specific for it. The present sequence
XX is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from USPTO at
XX seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
XX Sequence 458 BP; 103 A; 147 C; 123 G; 84 T; 0 U; 1 Other;
XX
XX Query Match 39.4%; Score 392; DB 8; Length 458;
XX Best Local Similarity 99.3%; Pred. No. 9.3e-85;
XX Matches 415; Conservative 0; Mismatches 0; Indels 3; Gaps 2;
XX
QY 580 TCCCGCCAGCAGTGGCGGACACCCCACTCCCTGCGCCAGTCCACAGTATTATTGT 639
XX
Db 17 TCCCGCCAGCAGTGGCGGACACCCCACTCCCTGCGCCAGTCCACAGTATTATTGT 76
XX
QY 640 TCCACATGGCTGGAGGGGGCCCTCCAGATTGGGGCCCTGGGTCCCACTCCCTG 699
XX
Db 77 TCCCAATATGGTGGAGGGGGCCCTCCAGATTGGGGCCCTGGGTCCCACTCCCTG 136
XX
QY 700 TCCATCCCGAGTGGGGCTGGACCCGCCAGATTCTCCCTTAAGGAATTGACATTCAGCAGG 759
XX

Db 137 TCCATCCCGAGTGGGGCTGGACCCGCCAGATTCTCCCTTAAGGAATTGACATTCAGCAGG 196
QY 760 GGTGGAGGCTCCACAGACCCAGGCGAGTGTGTGGAGGGGTGTTCCAAAGAGAGGCCT 819
Db 197 GGTGGAGGCTCCACAGACCCAGGCGAGTGTGTGGAGGGGTGTTCCAAAGAGAGGCCT 256
QY 820 GGTACAGACAGCGCCCGTGTCCCGCAGGTCTCGAGGCGACACTCGAGGGCCGAATTG 879
Db 257 GGTACAGACAGCGCCCGTGTCTCCCGCAGGTCTCGAGGCGACACTCGAGGGCCGAATTG 316
QY 880 TTCTTAGTTAGGCGCCAGCTCTCTGTTCAGTCCAAAGGTGAACACTCATGCGG--CAGC 937
Db 317 TTCTTAGTTAGGCGCCAGCTCTCTGTTCAGTCCAAAGGTGAACACTCATGCGGCGCAGC 376
QY 938 CATGGGCCCTCTGAGCAACTGTGTCAG-ACCCTTTCACCCCAATTAAACCCGAAACCA 994
Db 377 CATGGGCCCTCTGAGCAACTGTGTCAGCACCCTTTACCCCAATTAAACCCGAAACCA 434
XX
RESULT 10
ABAS2612/C
ID ABA62612 standard; DNA; 559 BP.
XX AC
XX ABA62612;
XX
DT 01-FEB-2002 (first entry)
XX
XX Human foetal liver single exon nucleic acid probe #10917.
XX
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
XX Homo sapiens.
XX
XX W0200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000669.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human foetal liver.
XX
XX Claim 1; SEQ ID NO 10917; 639pp + Sequence Listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human foetal liver. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 559 BP; 103 A; 176 C; 168 G; 112 T; 0 U; 0 Other;
XX
XX Query Match 21.9%; Score 217.6; DB 4; Length 559;
XX Best Local Similarity 96.1%; Pred. No. 1.1e-42;
XX Matches 223; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

QY      65 GCATGAGCGCGAGCTCAGGCGAGTGTAAGTCAACCAATCACTAACGCCAGCCAGT 124
Dd      544 GCCTGCCCTCCCTCCAGGCGAGTGTAAGTCAACCAATCACTAACGCCAGCCAGT 485

QY      125 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAAACGGGAGGGAGCTGCGCA 184
Dd      484 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAAACGGGAGGGAGCTGCGCA 425

QY      185 GGGTCCGCTGTGCTGCACCTGCTGTGTAAGCAGCAGTCAAGCCAGTCAAGCGGCGCTGCTCTGGC 244
Dd      424 GGGTCCGCTGTGCTGCACCTGCTGTGTAAGCAGCAGTCAAGCGGCGCTGCTCTGGC 365

QY      245 GGGTCCGCTGTGCTGCACCTGCTGTGTAAGCAGCAGTCAAGCGGCGCTGCTCTGGC 296
Dd      364 GGGTCCGCTGTGCTGCACCTGCTGTGTAAGCAGCAGTCAAGCGGCGCTGCTCTGGC 313

QY      245 GGCAGGAGAGATCAACCGGACCAAGGAGGAGCGCTGGAGCTGATCAACGG 296
Dd      364 GGCAGGAGAGATCAACCGGACCAAGGAGGAGCGCTGGAGCTGATCAACGG 313

RESULT 11
AAI42606/c
ID      AAI42606 standard; DNA; 559 BP.
XX
AC      AAI42606;
XX
DT      17-OCT-2001 (first entry)
XX
DE      Probe #11292 used to measure gene expression in human placenta sample.
XX
KW      Probe; microarray; human; placenta; antenatal diagnosis;
KW      genetic disorder; ss.
XX
OS      Homo sapiens.
XX
PN      WO200157272-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US000663.
XX
PR      04-FEB-2000; 2000US-0180312P.
PR      26-MAY-2000; 2000US-0207456P.
PR      30-JUN-2000; 2000US-00608408.
PR      03-AUG-2000; 2000US-00632366.
PR      21-SEP-2000; 2000US-0234687P.
PR      27-SEP-2000; 2000US-0236359P.
PR      04-OCT-2000; 2000GB-00024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-48897/53.
XX
XX
Human genome-derived single exon nucleic acid probes useful for analyzing
gene expression in human placenta.
XX
Claim 25; SEQ ID NO 11292; 654pp; English.
XX
The present invention relates to single exon nucleic acid probes (SENP).
The present sequence is one such probe. The probes are useful for
producing a microarray for predicting, measuring and displaying gene
expression in samples derived from human placenta. The probes are useful
for antenatal diagnosis of human genetic disorders
XX
Sequence 559 BP; 103 A; 176 C; 168 G; 112 T; 0 U; 0 Other;
Query Match      21.9%; Score 217.6; DB 4; Length 559;
Best Local Similarity 96.1%; Pred. No. 1.1e-42;
Matches 223; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      65 GCATGAGCGCGAGCTCAGGCGAGTGTAAGTCAACCAATCACTAACGCCAGCCAGT 124
Dd      544 GCCTGCCCTCCCTCCAGGCGAGTGTAAGTCAACCAATCACTAACGCCAGCCAGT 485

```

```

QY      125 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAAACGGGAGGGAGCTGCGCA 184
Dd      484 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAAACGGGAGGGAGCTGCGCA 425

QY      185 GGGTCCGCTGTGCTGCACCTGCTGTGTAAGCAGCAGTCAAGCCAGTCAAGCGGCGCTGCTCTGGC 244
Dd      424 GGGTCCGCTGTGCTGCACCTGCTGTGTAAGCAGCAGTCAAGCGGCGCTGCTCTGGC 365

QY      245 GGCAGGAGAGATCAACCGGACCAAGGAGGAGCGCTGGAGCTGATCAACGG 296
Dd      364 GGCAGGAGAGATCAACCGGACCAAGGAGGAGCGCTGGAGCTGATCAACGG 313

RESULT 12
AAK36818/c
ID      AAK36818 standard; DNA; 559 BP.
XX
AC      AAK36818;
XX
DT      06-NOV-2001 (first entry)
XX
DE      Human bone marrow expressed single exon probe SEQ ID NO: 11375.
XX
KW      Human; bone marrow expressed exon; gene expression analysis; probe;
KW      microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS      Homo sapiens.
XX
PN      WO200157272-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US000668.
XX
PR      04-FEB-2000; 2000US-0180312P.
PR      26-MAY-2000; 2000US-0207456P.
PR      30-JUN-2000; 2000US-00608408.
PR      03-AUG-2000; 2000US-00632366.
PR      21-SEP-2000; 2000US-0234687P.
PR      27-SEP-2000; 2000US-0236359P.
PR      04-OCT-2000; 2000GB-00024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-488900/53.
XX
XX
Human genome-derived single exon nucleic acid probes useful for analyzing
gene expression in human bone marrow.
XX
Example 4; SEQ ID NO 11375; 658pp + Sequence Listing; English.
XX
The present invention provides a number of single exon nucleic acid
probes which are derived from genomic sequences expressed in the human
bone marrow. They can be used to measure gene expression in bone marrow
samples, which may enable the improved diagnosis and treatment of cancers
such as lymphoma, leukaemia and myeloma. The present sequence is one of
the probes of the invention
XX
Sequence 559 BP; 103 A; 176 C; 168 G; 112 T; 0 U; 0 Other;
Query Match      21.9%; Score 217.6; DB 4; Length 559;
Best Local Similarity 96.1%; Pred. No. 1.1e-42;
Matches 223; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      65 GCATGAGCGCGAGCTCAGGCGAGTGTAAGTCAACCAATCACTAACGCCAGCCAGT 124
Dd      544 GCCTGCCCTCCCTCCAGGCGAGTGTAAGTCAACCAATCACTAACGCCAGCCAGT 485

QY      125 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAAACGGGAGGGAGCTGCGCA 184
Dd      484 GGGAGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAAACGGGAGGGAGCTGCGCA 425

```

```
QY 185 GGGTCCGCTGCTCGACCTGCTGTGGTGAAGCAGCCAGTACGGGGCCCTGCTCCTGGC 244
D 424 GGGTCCGCTGCTCGACCTGCTGTGGTGAAGCAGCCAGTACGGGGCCCTGCTCCTGGC 365
QY 245 GGCAGGAGAGATCACCCGGACCAAGGAGGAGGCCCTGGAGCTGATCAACGG 296
D 364 GGCAGGAGAGATCACCCGGACCAAGGAGGAGGCCCTGGAGCTGATCAACGG 313

RESULT 13
AAK10973/c
ID AAK10973 standard; DNA; 559 BP.
XX
AC AAK10973;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 10964.
XX
KW Human; brain expressed exon; gene expression analysis; probe; microarray;
KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer;
KW ss.
XX
OS Homo sapiens.
XX
FN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000667.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PS brains.
XX
PS Example 4; SEQ ID NO 10964; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention
XX
SQ Sequence 559 BP; 103 A; 176 C; 168 G; 112 T; 0 U; 0 Other;
Query Match 21.9%; Score 217.6; DB 4; Length 559;
Best Local Similarity 96.1%; Pred. No. 1.1e-42;
Matches 223; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 65 GCATGAGCGCAGCTCAGGCGGAGTGTACTTCAACCATCATCTAAGCCAGCCAGT 124
D 544 GCTTCCCTCCCTCCAGGCGGAGTGTACTTCAACCATCATCTAAGCCAGCCAGT 485
QY 125 GGGAGCGGCCAGCGGCAACAGCAGTGTGGTGGCAAAACGGGAGGGAGCTGGCA 184
D 484 GGGAGCGGCCAGCGGCAACAGCAGTGTGGTGGCAAAACGGGAGGGAGCTGGCA 425
```

```
QY 185 GGGTCCGCTGCTCGACCTGCTGTGGTGAAGCAGCCAGTACGGGGCCCTGCTCCTGGC 244
D 424 GGGTCCGCTGCTCGACCTGCTGTGGTGAAGCAGCCAGTACGGGGCCCTGCTCCTGGC 365
QY 245 GGCAGGAGAGATCACCCGGACCAAGGAGGAGGCCCTGGAGCTGATCAACGG 296
D 364 GGCAGGAGAGATCACCCGGACCAAGGAGGAGGCCCTGGAGCTGATCAACGG 313

RESULT 14
ABS10823/c
ID ABS10823 standard; DNA; 559 BP.
XX
AC ABS10823;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe from lung SEQ ID No 10814.
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
FN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US000665.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples.
XX
PS Claim 1; SEQ ID NO 10814; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of probes
CC; the novel set of probes which hybridise at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung, comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)
CC algorithmically predicting at least one exon from genomic sequences of
CC the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
```

CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridization to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene expression
CC analysis, and for identifying exons in a gene, particularly using human
CC lung derived mRNA and for the study of lung diseases such as asthma, lung
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC Karagenen syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC dyskinesia, pulmonary hypertension and biliary membrane disease. The
CC present sequence is a single exon probe of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 559 BP; 103 A; 176 C; 168 G; 112 T; 0 U; 0 Other;

Query Match 21.9%; Score 217.6; DB 6; Length 559;
Best Local Similarity 96.1%; Pred. No. 1.1e-42;
Matches 223; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 65 GCATGAGCGCGAGCTCAGCGCGAGTGTACTTCAACACATCACTAAGCGGCGAGCT 124
DB 544 GCGTCCCTCCCTCCAGGCGGAGTGTACTTCAACACATCACTAAGCGGCGAGCT 485
QY 125 GGGAGCGCGCGCGCGCGCGAGCGAGCGAGTGTGGGCAAAACGGGGAGCGCTGGCA 184
DB 484 GGGAGCGCGCGCGCGCGCGAGCGAGCGAGTGTGGGCAAAACGGGGAGCGCTGGCA 425
QY 185 GGGTCCGCTGCTCGACCTGCTGTTGAAGCAGCCAGTCAAGCGGCGCGCTGCTGGC 244
DB 424 GGGTCCGCTGCTCGACCTGCTGTTGAAGCAGCCAGTCAAGCGGCGCGCTGCTGGC 365
QY 245 GGCAGGAGAGATCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 296
DB 364 GGCAGGAGAGATCAACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 313

RESULT 15
ABA75131/c
ID ABA75131 standard; DNA; 213 BP.
XX ABA75131;
AC ABA75131;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #23436.
XX
KW Human, foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN W0200157277-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US000669.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.

XX Claim 4; SEQ ID NO 23436; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 213 BP; 25 A; 72 C; 67 G; 49 T; 0 U; 0 Other;

Query Match 21.4%; Score 213; DB 4; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.1e-41;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 83 GCGGAGTGTACTTCAACACATCACTAAGCGGCGAGTGGGAGCGGCGGCGGCA 142
DB 213 GCGGAGTGTACTTCAACACATCACTAAGCGGCGAGTGGGAGCGGCGGCGGCA 154
QY 143 ACAGCAGCAGTGGTGGCAAAACGGGGAGCGGCGGAGGAGGAGGAGGAGGAGGAGG 202
DB 153 ACAGCAGCAGTGGTGGCAAAACGGGGAGCGGCGGAGGAGGAGGAGGAGGAGGAGG 94
QY 203 TGCTGGTGAAGCAGCAGCGAGTCAAGCGGCGGCGCTGCTGGGCGGAGGAGGAGGAGG 262
DB 93 TGCTGGTGAAGCAGCAGCGAGTCAAGCGGCGGCGCTGCTGGGCGGAGGAGGAGGAGG 34
QY 263 GGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 295
DB 33 GGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1

Search completed: June 28, 2004, 16:36:42

Job time : 550.575 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2004, 14:35:25 ; Search time 113.078 Seconds
(without alignments)
4878.251 Million cell updates/sec

Title: U49070
Perfect score: 994
Sequence: 1 TGTGGCCAGCACCTCGAGG.....CCCAATTAAACCCAGAACCA 994

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgm2_6/ptodata/2/ina/5A COMB.seq:*
- 2: /cgm2_6/ptodata/2/ina/5B COMB.seq:*
- 3: /cgm2_6/ptodata/2/ina/6A COMB.seq:*
- 4: /cgm2_6/ptodata/2/ina/6B COMB.seq:*
- 5: /cgm2_6/ptodata/2/ina/PCTUS COMB.seq:*
- 6: /cgm2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	994	100.0	1014	2	US-09-066-074-1
2	994	100.0	1014	2	US-08-555-912A-1
3	994	100.0	1014	4	US-09-275-900-1
4	65.4	6.6	295	4	US-09-313-294A-5899
5	53.6	5.4	989	4	US-09-507-242-1
6	51.8	5.2	32998	4	US-09-408-020-1
7	49	4.9	42432	4	US-09-408-020-2
8	48.2	4.8	279	4	US-09-408-020-67
9	47.8	4.8	2097	4	US-09-252-991A-13635
10	47.8	4.8	2331	4	US-09-252-991A-13809
11	45.8	4.6	7218	1	US-08-232-463-14
12	45	4.5	1608	4	US-09-252-991A-11839
13	45	4.5	2640	4	US-09-252-991A-11547
14	45	4.5	2934	4	US-09-252-991A-11690
15	44.2	4.4	430	4	US-09-621-976-16656
16	44.2	4.4	1266	4	US-09-252-991A-2521
17	44.2	4.4	1377	4	US-09-252-991A-2310
18	44.2	4.4	1428	4	US-09-252-991A-2606
19	43	4.3	1260	4	US-09-252-991A-14282
20	43	4.3	2175	4	US-09-252-991A-14313
21	43	4.3	2244	4	US-09-252-991A-14391
22	42	4.2	446	4	US-09-621-976-17380
23	42	4.2	480	4	US-09-621-976-17379
24	42	4.2	497	4	US-09-621-976-17378
25	42	4.2	605	3	US-09-208-804-2
26	42	4.2	605	3	US-08-801-743-2
27	42	4.2	605	4	US-09-016-434-270

C 28 42 4.2 1029 4 US-09-252-991A-13686
 29 42 4.2 1071 4 US-09-252-991A-13459
 30 42 4.2 1809 4 US-09-252-991A-13887
 31 41 4.1 741 4 US-09-489-039A-3673
 32 40.8 4.1 347 4 US-09-636-215-725
 33 40.8 4.1 347 4 US-09-685-166A-725
 34 40.8 4.1 1594 4 US-09-566-921-26
 35 39.8 4.0 44377 2 US-08-804-227C-7
 36 39.8 4.0 44377 2 US-08-804-198-1
 37 39.6 4.0 333 4 US-09-489-039A-859
 38 39.4 4.0 1167 1 US-07-960-985-1
 39 39.4 4.0 1167 2 US-08-496-671-1
 40 39.4 4.0 4080 4 US-09-016-434-1342
 41 39.2 3.9 336 4 US-09-252-991A-3451
 42 39.2 3.9 1218 4 US-09-252-991A-3440
 43 39.2 3.9 77536 4 US-09-410-551B-1
 44 39 3.9 3468 1 US-07-951-715A-2
 45 39 3.9 3468 1 US-07-951-715A-4

ALIGNMENTS

RESULT 1
 US-09-066-074-1
 ; Sequence 1, Application US/09066074
 ; Patent No. 5952467
 ; GENERAL INFORMATION:
 ; APPLICANT: Hunter et al., Tony
 ; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C.
 ; STREET: 4225 Executive Square, Suite 1400
 ; CITY: La Jolla
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 92037
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/066,074
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/555,912
 ; FILING DATE: 13-NOV-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Haile, Lisa A.
 ; REGISTRATION NUMBER: 38,347
 ; REFERENCE/DOCKET NUMBER: 07251/011001
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 619/678-5070
 ; TELEFAX: 619/678-5099
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1014 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 25..513
 ; US-09-066-074-1

Query Match 100.0%; Score 994; DB 2; Length 1014;
 Best Local Similarity 100.0%; Pred. No. 6.2e-224;
 Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGGCCAGCACCTCGAGGAGAGATGCGGACGAGGAGAGCTGCGCGCGGTGGAG 60

RESULT 2
US-08-555-912A-1
; Sequence 1, Application US/08555912A
; Patent No. 597697
; GENERAL INFORMATION:

Query Match	100.0%;	Score 994;	DB 2;	Length 1014;
Best Local Similarity	100.0%;	Pred. No. 6.2e-224;		
Matches 994;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	TGCTGGCCAGCACTTCGAGGGAAGATGGCGGACGAGGAGAACTCGCGCCGGCTGGGAG	60	
DB	1	TGCTGGCCAGCACTTCGAGGGAAGATGGCGGACGAGGAGAACTCGCGCCGGCTGGGAG	60	
QY	61	AAGCGCATGAGCGCGAGCTCAGCGCGAGTGTACTTCTCAACCAATCACTAAACGCCAGC	120	
DB	61	AAGCGCATGAGCGCGAGCTCAGCGCGAGTGTACTTCTCAACCAATCACTAAACGCCAGC	120	
QY	121	CAGTGGGACGGGCCACAGCGGCAACAGCAGCAGTGTGTGGCAAAAACGGGACGGGGAGCCT	180	
DB	121	CAGTGGGACGGGCCACAGCGGCAACAGCAGCAGTGTGTGGCAAAAACGGGACGGGGAGCCT	180	
QY	181	GCCAGGGTCGGTGTCTCGCACCTGCTGTGTGAGACACAGCCAGTTCAGCGGCGCCCTCGTCC	240	
DB	181	GCCAGGGTCGGTGTCTCGCACCTGCTGTGTGAGACACAGCCAGTTCAGCGGCGCCCTCGTCC	240	
QY	241	TGCGGCGACGAGAAAGATCACCCGGACCAAGGAGGAGGCGCCTTGAGCTGTATCAACGGCTAC	300	
DB	241	TGCGGCGACGAGAAAGATCACCCGGACCAAGGAGGAGGCGCCTTGAGCTGTATCAACGGCTAC	300	
QY	301	ATCCAGAAGATCAAGTCGGGAGAGGAGCACTTTGAGTCTCTGGCCCTCACAGTTACGGCAC	360	
DB	301	ATCCAGNAGATCAAGTCGGGAGAGGAGCACTTTGAGTCTCTGGCCCTCACAGTTACGGCAC	360	
QY	361	TGACGTTCAGCCAAAGCCAGGGGAGACCTGGGTGCCTTCAGCAGAGGTTCAGATGCAGAAAG	420	
DB	361	TGACGTTCAGCCAAAGCCAGGGGAGACCTGGGTGCCTTCAGCAGAGGTTCAGATGCAGAAAG	420	
QY	421	CCATTTGAAGACGCCTGTTTGCCTTCGGACCGGGGAGATGATGCGGGGCCGTGTTTCACG	480	
DB	421	CCATTTGAAGACGCCTGTTTGCCTTCGGACCGGGGAGATGATGCGGGGCCGTGTTTCACG	480	

481 GATTCGGGATCCACATCATCTCCGCACTGAGTGAGGAGCCAGCCCTGGCT 540
Db
481 GATTCGGGATCCACATCATCTCCGCACTGAGTGAGGAGCCAGCCCTGGCT 540
Qy
541 CGGGCAGGGAGGGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGG 600
Db
541 CGGGCAGGGAGGGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGGAGGGG 600
Qy
601 ACCCCCACTCCCTCCACCGCTCACAGATATTATTGTTCCACAATGGCTGGAGGGG 660
Db
601 ACCCCCACTCCCTCCACCGCTCACAGATATTATTGTTCCACAATGGCTGGAGGGG 660
Qy
661 GCCCTTCCAGATTGGGGGCGCTGGGGTCCCACTCCCTGTCCATCCCAAGTTGGGCTGC 720
Db
661 GCCCTTCCAGATTGGGGGCGCTGGGGTCCCACTCCCTGTCCATCCCAAGTTGGGCTGC 720
Qy
721 GACCGCAGATTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGAGGCTCCACAGACCA 780
Db
721 GACCGCAGATTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGAGGCTCCACAGACCA 780
Qy
781 GGGCAGTGTGGGAGGGGTGTTCCAAAGAGAGGCTGGTTCAGCAGAGCCGCCCGTG 840
Db
781 GGGCAGTGTGGGAGGGGTGTTCCAAAGAGAGGCTGGTTCAGCAGAGCCGCCCGTG 840
Qy
841 TCCCCCAAGTGTGGAGGAGCAGCTCAGAGGCGGAATTGTTCTAGTTAGGCGACGCTCC 900
Db
841 TCCCCCAAGTGTGGAGGAGCAGCTCAGAGGCGGAATTGTTCTAGTTAGGCGACGCTCC 900
Qy
901 TCTGTTGAGTGCAGAGGTGAACACTCATGCGGAGGAGGCTGAGCAACTGTG 960
Db
901 TCTGTTGAGTGCAGAGGTGAACACTCATGCGGAGGAGGCTGAGCAACTGTG 960
Qy
961 CAGACCTTTTCAACCCCAATTAACCCAGAACCA 994
Db
961 CAGACCTTTTCAACCCCAATTAACCCAGAACCA 994

RESULT 3

US-09-275-900-1
; Sequence 1, Application US/09275900
; Patent No. 6596848
; GENERAL INFORMATION:
; APPLICANT: Hunter et al., Tony
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/275,900
; FILING DATE: 24-Mar-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/555,912
; FILING DATE: 13-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07251/011001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

LENGTH: 1014 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 25..513
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-275-900-1

Query Match 100.0%; Score 994; DB 4; Length 1014;
Best Local Similarity 100.0%; Pred. No. 6.2e-224;
Matches 994; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 TGTGCGCAGCACTTCGAGGGAAGATGGCGGAGGAGAACTGCGCGCGCTGGAG 60
Db 1 TGTGCGCAGCACTTCGAGGGAAGATGGCGGAGGAGAACTGCGCGCGCTGGAG 60
Qy 61 AAGCGCATGAGCGCGAGCTCAGCGCGAGTGTACTTCAACACATCACTAACGCCAG 120
Db 61 AAGCGCATGAGCGCGAGCTCAGCGCGAGTGTACTTCAACACATCACTAACGCCAG 120
Qy 121 CAGTGGGAGCGGCCAGCGGCAACAGCAGAGTGTGGCAAAACCGGCGAGGGAGCCT 180
Db 121 CAGTGGGAGCGGCCAGCGGCAACAGCAGAGTGTGGCAAAACCGGCGAGGGAGCCT 180
Qy 181 GCCAGGTCGGTGTGTCGACCTGCTGAGCAGAGCAGCGAGTGTGGCGGCGCTTCT 240
Db 181 GCCAGGTCGGTGTGTCGACCTGCTGAGCAGAGCAGCGAGTGTGGCGGCGCTTCT 240
Qy 241 TGGCGGCGAGGAGAGATCACCCGAGCAAGAGGAGGAGCCCTGGAGCTGATCAAC 300
Db 241 TGGCGGCGAGGAGAGATCACCCGAGCAAGAGGAGGAGCCCTGGAGCTGATCAAC 300
Qy 301 ATCCAGAGATCAAGTGGGAGAGGAGCACTTTGAGTCTCTGGCTTCAAGTTGAG 360
Db 301 ATCCAGAGATCAAGTGGGAGAGGAGCACTTTGAGTCTCTGGCTTCAAGTTGAG 360
Qy 361 TGCAGCTCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420
Db 361 TGCAGCTCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 420
Qy 421 CCATTTGAAGACGCTCTGTTGGCTGCGGAGCGGAGGAGGAGGAGGAGGAGGAG 480
Db 421 CCATTTGAAGACGCTCTGTTGGCTGCGGAGCGGAGGAGGAGGAGGAGGAGG 480
Qy 481 GATTCCGGCATCCACATCATCTCCGCACTGAGTGTGGGAGGAGGAGGAGGAGG 540
Db 481 GATTCCGGCATCCACATCATCTCCGCACTGAGTGTGGGAGGAGGAGGAGGAGG 540
Qy 541 CGGGCAGGGCAGGGCGGCTAGGCGCGAGTCCCTTCCCGCGGAGGAGGAGGAGG 600
Db 541 CGGGCAGGGCAGGGCGGCTAGGCGCGAGTCCCTTCCCGCGGAGGAGGAGGAGG 600
Qy 601 ACCCCCACTCCCTGCCACCGCTCACAGATATTATTGTTCCACAATGGCTGGAG 660
Db 601 ACCCCCACTCCCTGCCACCGCTCACAGATATTATTGTTCCACAATGGCTGGAG 660
Qy 661 GCCCTTCCAGATTGGGGGCGCTGGGGTCCCACTCCCTGTCCATCCCAAGTTGG 720
Db 661 GCCCTTCCAGATTGGGGGCGCTGGGGTCCCACTCCCTGTCCATCCCAAGTTGG 720
Qy 721 GACCGCAGATTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGAGGCTCCACAG 780
Db 721 GACCGCAGATTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGAGGCTCCACAG 780
Qy 781 GGGCAGTGTGGGAGGGGTGTTCCAAAGAGAGGCTGGTTCAGCAGAGCCGCCCG 840
Db 781 GGGCAGTGTGGGAGGGGTGTTCCAAAGAGAGGCTGGTTCAGCAGAGCCGCCCG 840
Qy 841 TCCCGCCAGGTCGTGGAGGAGCAGACTCAGAGGCGGAATTGTTCTAGTTAGGCG 900
Db 841 TCCCGCCAGGTCGTGGAGGAGCAGACTCAGAGGCGGAATTGTTCTAGTTAGGCG 900

QY 901 TCTGTTTCAGTCGCAAAAGTGAACACTCATCGCGGAGGCATGGCCCTCTGAGCAACTGTG 960
Db 901 TCTGTTTCAGTCGCAAAAGTGAACACTCATCGCGGAGGCATGGCCCTCTGAGCAACTGTG 960
QY 961 CAGACCCCTTTCACCCCAATTAACCCAGAACCA 994
Db 961 CAGACCCCTTTCACCCCAATTAACCCAGAACCA 994

RESULT 4

US-09-313-294A-5899
; Sequence 5899, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 5899
; LENGTH: 295
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700350866H1
; NAME/KEY: unsure
; LOCATION: 244, 247, 252-255, 260-261, 276, 291
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-5899

Query Match 6.6%; Score 65.4; DB 4; Length 295;
Best Local Similarity 66.9%; Pred. No. 1.5e-06;
Matches 93; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
QY 385 GACCTGGTGCCCTTCAGCAGAGTTCAGATCGAAGCCCAATTTGAAGCGCCTCGTTGGC 444
Db 1 GACTTAGTACCTTTGGGAGGAGCGAGATGCAAAACCCCTTCGAGCAGCCCATATGCC 60
QY 445 CTCGCGAGCGGAGATGAGCGGCCCGCTTTCAGCGATTCCGGCATCCACATCATCCTC 504
Db 61 CTCAGGTTGGTGAGCTCAGCGACATCGTGACACTGACAGTGGGGTTACATCATCCTG 120
QY 505 CGCACTGAGTGAGGTGGG 523
Db 121 CGGACTGCTGAGGAGGG 139

RESULT 5

US-09-507-242-1
; Sequence 1, Application US/09507242
; Patent No. 6537753
; GENERAL INFORMATION:
; APPLICANT: HANES, Steven D.
; APPLICANT: DEVASAHAYAM, Gina
; APPLICANT: CHATURVEDI, Vishnu
; TITLE OF INVENTION: CABSSI: A CANDIDA ALBICANS GENE, METHODS FOR MAKING AND
; TITLE OF INVENTION: USING, AND TARGETING IT OR ITS EXPRESSION PRODUCTS FOR
; FILE REFERENCE: 454311-2200.1
; CURRENT APPLICATION NUMBER: US/09/507,242
; CURRENT FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 60/121,246
; PRIOR FILING DATE: 1999-02-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 989

TYPE: DNA
; ORGANISM: Candida albicans
US-09-507-242-1
Query Match 5.4%; Score 53.6; DB 4; Length 989;
Best Local Similarity 48.4%; Pred. No. 0.0011;
Matches 180; Conservative 0; Mismatches 189; Indels 3; Gaps 1;
QY 148 AGCAGTGTGGCAAAACAGCGGAGCGGAGCCTGCGCAGGCTCCGCTGCTGCACTGCTG 207
Db 418 AACAATGGTTACAGCCACTTGTGATGAGATGGCCAGGTAGAGTTTCTCATTTGTTG 477
QY 208 GTGAAGCACAGCCAGTCCAGCGGCCCTCGCTCTGCGGG---CAGGAGAGATCACCCGG 264
Db 478 ATCAAGAACATCAATCAAGAAACCCCAAGTCTTTGGAAGTCCCAGATGGTATAAGTAGA 537
QY 265 ACCAAGGAGGAGGCCCTGGAGCTGATCAACGGGTACATCCAGAGATCAAGTCCGGAGAG 324
Db 538 ACTAGAGACGAATCTATACAGATATTGAAGAAACATTTGGAAGAAATATTGAGTGGTGG 597
QY 325 GAGGACTTTGAGTCTCTGGCCCTCACAGTTCAGCGACTGCAGCTCAGCCAGCCAGCGGGA 384
Db 598 GTTAACTAAGTGAATTGGCAATACCGAAAGTGTTCAGCTCACATGACAGAGTGGT 657
QY 385 GACCTGGTGCCCTTCAGCAGAGGTGAGATGAGAGCCATTTGAAGCGCCTCGTTTCGG 444
Db 658 GATTTAGGGTTTTTTAGCAAAAGGACAAATGCAACCCATTCGAAGAAAGCGCATTCAT 717
QY 445 CTGCGGACGGGAGATGAGCGGCCCGCTGTTACGGATTCCGGCATCCACATCATCTC 504
Db 718 TTGCATGTTGAGAGAGTCAGTAACAATAATTGAAACCAATAGTGGTGTCCATATCTCCA 777
QY 505 CGCACTGAGTGA 516
Db 778 AGAACAGGATTA 789

RESULT 6

US-09-408-020-1/c
; Sequence 1, Application US/09408020
; Patent No. 6632937
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOPT.002A
; CURRENT APPLICATION NUMBER: US/09/408,020
; CURRENT FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 32998
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7604)...(8908)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8961)...(9767)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10545)...(10922)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (13944)...(14612)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (18638)...(20149)
; FEATURE:

```
NAME/KEY: CDS
LOCATION: (20554)....(20955)
FEATURE:
NAME/KEY: CDS
LOCATION: (20956)....(21834)
FEATURE:
NAME/KEY: CDS
LOCATION: (25151)....(26377)
FEATURE:
NAME/KEY: CDS
LOCATION: (27535)....(28002)
FEATURE:
NAME/KEY: CDS
LOCATION: (28065)....(29483)
US-09-408-020-1

Query Match
Score 51.8; DB 4; Length 32998;
Best Local Similarity 48.2%; Pred. No. 0.0064;
Matches 146; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 311 TCAAGTCGGGAGAGGAGCTTTGAGTCTCTGGCTCAGCTTCAGGACTGCAGCTCAG 370
DB 20361 TCAAGCGGGGAAAGTTTGAAAGCTGGCAAGAGCTCTCGATAGACGGGGGCGCG 20302

QY 371 CCAAGCGCGGAGAGCTGGTGGCTTCAGCAGAGTTCAGATCAGAGCCATTGAG 430
DB 20301 CAAAGAGGAGCGGAGCTTGGGCTACTTTGGCAGGGGCAAGATGTTAAGCCGTTGAGG 20242

QY 431 ACGCTGTTTGGCTGGGAGCGGGGAGATGAGCGGGCCGCTGTTACGGATTCGGCA 490
DB 20241 ATGCGCGGCTCCGCTCGAGTGGAGGATTCGAGCGGTAAGTCCGAGTTGCT 20182

QY 491 TCCACATCATCTCCGACTCAGTGAGGTTGGGAGCCAGCCGCTGGCTCGGGGCGAGG 550
DB 20181 ACCAGCTAAGCGCTGGGATAAGCCGCTAGAAATAGCCCTCGAGCCCTTTT 20122

QY 551 CAGGCGGCTAGGCGGCGAGCTCCCTTGGCCCGCAGCAGTGGCGGACCCCGGACT 610
DB 20121 GGGGCTGCTTTGGGCGAGGGGCGGCGGCTGCAAGCGACTTGTTCATCTGTCCCGCATG 20062

QY 611 CCC 613
DB 20061 CCC 20059

RESULT 7
US-09-408-020-2/c
Sequence 2, Application US/09408020
Patent No. 6632937
GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.
APPLICANT: Schleper, Christa
TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
FILE REFERENCE: DCOIP 002A
CURRENT APPLICATION NUMBER: US/09/408,020
CURRENT FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: 60/102,294
PRIOR FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 123
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 42432
TYPE: DNA
ORGANISM: Cenarchaeum symbiosum
FEATURE:
NAME/KEY: CDS
LOCATION: (3)....(10421)
FEATURE:
NAME/KEY: CDS
LOCATION: (10625)....(11434)
FEATURE:
NAME/KEY: CDS
```

```
LOCATION: (11478)....(13046)
FEATURE:
NAME/KEY: CDS
LOCATION: (13046)....(14620)
FEATURE:
NAME/KEY: CDS
LOCATION: (23558)....(24862)
FEATURE:
NAME/KEY: CDS
LOCATION: (24913)....(25728)
FEATURE:
NAME/KEY: CDS
LOCATION: (26504)....(26881)
FEATURE:
NAME/KEY: CDS
LOCATION: (29655)....(30491)
FEATURE:
NAME/KEY: CDS
LOCATION: (34559)....(36067)
FEATURE:
NAME/KEY: CDS
LOCATION: (37002)....(37403)
FEATURE:
NAME/KEY: CDS
LOCATION: (37404)....(38282)
FEATURE:
NAME/KEY: CDS
LOCATION: (39454)....(40572)
US-09-408-020-2

Query Match
Score 4.9%; DB 4; Length 42432;
Best Local Similarity 55.6%; Pred. No. 0.03;
Matches 94; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 352 TTCAGCGACTCAGCTAGCCAGGCGGAGACCTGGGTGCTTCAGCAGAGGTGAG 411
DB 36228 TCCATGGACGGGGGAGTGCAGAGGAGGACGCGAGTCTCGGCTACTTTGGCAGGGGAAAG 36169

QY 412 ATGCAGAGCCATTGGAAGACGCTTCGTTTCGCTGCGGAGCGGGGAGATGAGCGGGCCC 471
DB 36168 ATGTTAAAGCCGTTTGAAGAGCGCGCTTCGCGCTGCAGATAGCGAGATCTCGGAACCC 36109

QY 472 GTGTTACGGATTCGGGATCCACATCATCTCCGCACTGAGTGAGGT 520
DB 36108 ATAAAGTCAGAGTTTGGCTATCACGTGATAAAGCGCTCGGCTAAAGT 36060

RESULT 8
US-09-408-020-67
Sequence 67, Application US/09408020
Patent No. 6632937
GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.
APPLICANT: Schleper, Christa
TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
FILE REFERENCE: DCOIP 002A
CURRENT APPLICATION NUMBER: US/09/408,020
CURRENT FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: 60/102,294
PRIOR FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 123
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 67
LENGTH: 279
TYPE: DNA
ORGANISM: Cenarchaeum symbiosum
FEATURE:
NAME/KEY: CDS
LOCATION: (1)....(279)
US-09-408-020-67

Query Match
Score 4.8%; DB 4; Length 279;
```

Tue Jun 29 08:37:22 2004

Best Local Similarity 52.8%; Pred. No. 0.016;
Matches 104; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 311 TCAAGTCGGAGAGGAGGACTTTGAGTCTCTGGCTCAGCTTCCAGGACTGCGAGCTCAG 370
Db 74 TCAGGCGGGGGAAGAGTTTGAAGCTGCGCAAGGAGCTCTCGATAGCGGGGAGCG 133

QY 371 CCAAGCGCAGGGAGAGCTGGGTGCTTTCAGAGAGTTCAGATCAGAGGAGCATTGAG 430
Db 134 CAAAGAGGAGCGGAGCTTGGGCTACTTTGGCAGGGCAGATGCTAAAGCGGTTGAGG 193

QY 431 AGCCTCTGTTGCGTGGCGGAGAGTGGAGGCGGCGGTGTTACCGATTCCGGCA 490
Db 194 ATGCGCGGTTCGCTGCGAGTAGGAGGATTCGAGCGGTAATCCGAGTTGGCT 253

QY 491 TCCACATCATCTCCGC 507
Db 254 ACCAGTGATAAGCGC 270

RESULT 9

US-09-252-991A-13635
Sequence 13635, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 13635

LENGTH: 2097

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-13635

Query Match 4.8%; Score 47.8; DB 4; Length 2097;

Best Local Similarity 49.8%; Pred. No. 0.03;

Matches 149; Conservative 0; Mismatches 147; Indels 3; Gaps 1;

QY 206 TGGTGAAGCACAGCCAGTCAAGCGGCGCTCTGCTGCGGCGAGAGATCACCCGGA 265

Db 1004 TCGCCAACTCTCCGAGCAGCGGATCGCCGCCACATCTCTGATCGAGTGAACGACAAGG 1063

QY 266 CCAAGGAGGAGGCGCTGAGTCAACGCTACATCCAGAGATCAAGTCCGGAGAGG 325

Db 1064 TCGCGCAGCAGCAGGCGCAAGCGGAGATCAGCAGAGATCAAGCTCGCTGCGCAGGCGG 1123

QY 326 AGGACTTTGAGTCTCTGCTCTCAGTTTCAGC---GACTGCGAGCTCAGCCAGCCAGG 382

Db 1124 AGGATTTGCGCGCTGGCCAGAGTTCTCCAGGATATCGGCTCGCGCCAGCCAGCGG 1183

QY 383 GAGACCTGGGTGCTTTCAGCAGAGGTCAGATCGAGAGCCATTGAGAGCGCTCTGTTG 442

Db 1184 GCGACCTGGGTGCTTTCAGCAGAGGTCAGATCGAGAGCCATTGAGAGCGCTCTGTTG 1243

QY 443 CGCTCGGAGCGGGAGATGAGCGGCGCGGTTTCAGGATTCGGGATCCACATCATC 501

Db 1244 CGCTGAAGCAAGGTGAGGTATCCGCGCGCGGTGAAGACTCCGTCAGGCTACCACTGATC 1302

RESULT 10

US-09-252-991A-13809/C

Sequence 13809, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 13809

LENGTH: 2331

TYPE: DNA

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-13809

Query Match 4.8%; Score 47.8; DB 4; Length 2331;

Best Local Similarity 49.8%; Pred. No. 0.031;

Matches 149; Conservative 0; Mismatches 147; Indels 3; Gaps 1;

QY 206 TGGTGAAGCACAGCCAGTCAAGCGGCGCTCTGCTGCGGCGAGAGATCACCCGGA 265

Db 1505 TCGCCAACTCTCCGAGCAGCGGATCGCCGCCACATCTCTGATCGAGTGAACGACAAGG 1446

QY 266 CCAAGGAGGAGGCGCTGAGTCAACGCTACATCCAGAGATCAAGTCCGGAGAGG 325

Db 1445 TCGCGCAGCAGCAGGCGCAAGCGGAGATCGAGAGATCAAGGCTCGCTGCGCCAGGCGG 1386

QY 326 AGGACTTTGAGTCTCTGCTCTCAGTTTCAGC---GACTGCGAGCTCAGCCAGGCGCAGG 382

Db 1385 AGGATTTGCGCGCTGCGCCAGGAGTTCTCCAGGATATCGGCTCGCGCCAGCCAGGCGG 1326

QY 383 GAGACCTGGGTGCTTTCAGCAGAGTCAAGTCAAGAGCCATTGAGAGCGCTCTGTTG 442

Db 1325 GCGACTTGGGCTACCGCGCGCGCTGTACGACCCGCGTTTCAGGAGGCGGCTGTATG 1266

QY 443 CGCTCGGAGCGGGAGATGAGCGGCGCGGTTTCAGGATTCGGGATTCGGGATCCACATCATC 501

Db 1265 CGCTGAAGCAAGGTGAGGTATCCGCGCGCGGTGAAGACTCCGTCAGGCTACCACTGATC 1207

RESULT 11

US-08-232-463-14/c

Sequence 14, Application US/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.

APPLICANT: SCHEIFELINGER, F.

APPLICANT: FALKNER, F. G.

TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 1800 Diagonal Road, Suite 500

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,463

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/935,313

FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 4.6%; Score 45.8; DB 1; Length 7218;
Best Local Similarity 4.9%; Pred. No. 0.12;
Matches 20; Conservative 216; Mismatches 173; Indels 0; Gaps 0;
QY 17 GAGGGAAGATGGCGGACGAGGAGAGCTCCGCCCGGCTGGGAGAGAGCCATGAGCCGCA 76
Db 1433 RRR 1374
QY 77 GTCAGCGCGAGTGTTACTTCAACCATCATCTAACCCAGCCAGCTGGGAGCGGCCCA 136
Db 1373 RRR 1314
QY 137 GCGCAACAGCAGCTGGTGCAAAACGGGAGGAGCGCTGCGAGGTCGCTGCTGCT 196
Db 1313 RRR 1254
QY 197 CGCACTGCTGGTGAAGACAGCAGCAGCTGCGCGGCGCTGCTGCTGCGGAGAGAGA 256
Db 1253 RRR 1194
QY 257 TCACCGGACAGGAGGAGCGCTGAGCTGATCAACGCTACATCCAGAGTCAAGT 316
Db 1193 RRR 1134
QY 317 CGGAGAGGAGGAGCTTTGAGTCTCTGCGCTCACAGTTGAGCTGAGCTGAGCCAAAG 376
Db 1133 RRR 1074
QY 377 CCAGGGGAGACCTGGTGCTTACGAGAGAGTGCAGTGCAGAGCCATT 425
Db 1073 RRRRRRRATCGAAGCTCCCTCGACCTGACCCAAAGCTCGGAATTAATT 1025

RESULT 12
US-09-252-991A-11839
Sequence 11839, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 11839
LENGTH: 1608
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11839
Query Match 4.5%; Score 45; DB 4; Length 1608;

Best Local Similarity 47.7%; Pred. No. 0.13;
Matches 132; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
QY 52 GGCTGGGAGAGCGCATGAGCCGAGCTCAGCGCGAGTGTACTTCAACCATCACT 111
Db 953 GTCTGGGACATGCTGCTGCGCGCAGGAGCCACCTGGCGGACTACGACCCGGGACC 1012
QY 112 AACGCCAGCCAGTGGGAGCGGCCAGCGGCAACAGCAGTGTGGTGGCAAAACGGGCG 171
Db 1013 AAGGCGAGCATTTGGCGAGCTGCGACCGCGCTGCTGTTCAATCCCGAGCGCGGCGG 1072
QY 172 GGGAGCGCTGCGAGTCCGCTGCTCGCACCTGCTGTGTAAGCAGCAGCCAGTCAAGGG 231
Db 1073 GTCAAGCATTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1132
QY 232 CCTCGTCTGCGCGCAGGAGAGATCACCCGAGCAACCGGAGAGGAGGCGCTGGAGTGT 291
Db 1133 GCGTCTGCGCTGCGCGAGGAGCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1192
QY 292 AACGCTCATCTCAGAGATCAAGTTCGGGAGGAGG 328
Db 1193 TACGAGCAGCCCAACATGCTCAACACCGAGAACCCAGG 1229

RESULT 13
US-09-252-991A-11547/c
Sequence 11547, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 11547
LENGTH: 2640
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11547
Query Match 4.5%; Score 45; DB 4; Length 2640;
Best Local Similarity 47.7%; Pred. No. 0.14;
Matches 132; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
QY 52 GGCTGGGAGAGCGCATGAGCCGAGCTCAGCGCGAGTGTACTTCAACCATCACT 111
Db 1818 GTCTGGGACATGCTGCTGCGCGCAGGAGCGCCACCTGGCGGACTACGACCCGGGACC 1759
QY 112 AACGCCAGCGAGTGGGAGCGGCCAGCGGCAACAGCAGTGTGGTGGCAAAACGGGCG 171
Db 1758 AAGGCGAGCATTTGGCGAGCTGCGACCGCGCTGCTGTTCAATCCCGAGCGCGGCGG 1699
QY 172 GGGAGCGCTGCGAGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 231
Db 1698 GTCAAGCATTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1639
QY 232 CCTCGTCTGCGCGCAGGAGAGATCACCCGAGCAACCGGAGGAGGCGCTGGAGTGT 291
Db 1638 GCGTCTGCGCTGCGCGAGGAGCCAGCGCTGAGTTCAGTTCAGTTCAGTTCAGTTC 1579
QY 292 AACGCTCATCTCAGAGATCAAGTTCGGGAGGAGG 328
Db 1578 TACGAGCAGCCCAACATGCTCAACACCGAGAACCCAGG 1542

RESULT 14
US-09-252-991A-11690
Query Match 4.5%; Score 45; DB 4; Length 1608;

; Sequence 11690, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11690
; LENGTH: 2934
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11690

Query Match 4.5%; Score 45; DB 4; Length 2934;
Best Local Similarity 47.7%; Pred. No. 0.15;
Matches 132; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
QY 52 GGCCTGGAGAGCGCATGAGCGCGAGCTCAGCGCGAGTGTACTTCAACACATCACT 111
DB 1174 GTCTGGGACATGCTGGTCCGCCGCGAGCGGACCTGGGCGACTACGACCCGGGGACC 1233
QY 112 AAGCCAGCAGCAGTGGGAGCGGCCAGCGGCAACAGCAGCAGTGGTGGCAAAAACGGGCGAG 171
DB 1234 AAGGGCAGCATTTGGCGAGCTGGCGACCCGGCGCTGGTTCAATCCCGAGGCGGGCGAGCGG 1293
QY 172 GGGGAGCCTGCCAGGCTCCGCTGCTCGACCTGCTGGTGAAGCAGACCCAGTCAACGGCGG 231
DB 1294 GTCAAGCATTTGGCGGCTGGGCTACTCGGCTACGTGATGCGCTCGGACTGGCCAAAGCTT 1353
QY 232 CCCTGCTCTGGCGGCGAGGAGATACCCGACCAAGGAGGAGCGCCCTGGAGCTGATC 291
DB 1354 GGGCTGCGCCCTGGCGAGGACGACGCGCTGATGCTAGTACCTGACCCAGGTGTCC 1413
QY 292 AAGCGCTACATCCAGAGATCAAGTCCGGAGAGGAGG 328
DB 1414 TAGGACGACGCAACATGCTCAACACCGAGAACGAGG 1450

RESULT 15
US-09-621-976-16656
; Sequence 16656, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621.976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 16656
; LENGTH: 430
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-16656

Query Match 4.4%; Score 44.2; DB 4; Length 430;
Best Local Similarity 18.9%; Pred. No. 0.15;
Matches 49; Conservative 101; Mismatches 109; Indels 0; Gaps 0;
QY 156 TGGCAAAACGGGAGCGGAGCCTGCCAGGTCGCTCGCACCTGCTGGTGAAGCA 215
DB 31 TGGCMMCMRSCCYMMCGGRRSCGRAMCCYYKKXGSCRAMCCCCYTYCKSCSS 90

QY 216 CAGCCAGTCACGGCGCCCTCTGCTCGCGCAGGAGAAGATCACCCGACCAAGAGGA 275
DB 91 YKGYTTKRAMMKRRSCYTYRRRMYYYWRSYMMRSMKSGSCCCGSGGSCYKKKKKGK 150
QY 276 GGCCTGGAGTGTATCAACGGCTACATCCAGAAGATCAAGTCGGGAGAGGACTTTGA 335
DB 151 KGSCCMRSYWCCYYKRAARMKMMKGGSCMYYTKRMWRMCCCCMRRRRRMRMRMCMWK 210
QY 336 GTCTCTGGCCTCACAGTTTCAGCGACTGCGACTCAGCCCAAGGCCAGGGGAGACCTGGGTGC 395
DB 211 SYTCYKSSSMCMARWRKARKKRMCCYTKGGGRMRKYCOMRKKGRACCTGTTTAC 270
QY 396 CTTCAGCAGAGGTGATG 414
DB 271 CTGCCGTGCTGTCGAGAG 289
Search completed: June 28, 2004, 18:04:34
Job time : 115.078 secs


```
RESULT 2
US-10-108-260A-4441
; Sequence 4441, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4441
; LENGTH: 145
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-4441

Query Match      78.1%; Score 665; DB 15; Length 145;
Best Local Similarity 98.4%; Pred. No. 5.2e-61;
Matches 126; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MADEKLPGWKMSRSGRVYFNHITNASQWERPSNGSGNGGQEPARVRCSHL 60
Db 1 MADEKLPGWKMSRSGRVYFNHITNASQWERPSNGSGNGGQEPARVRCSHL 60

QY 61 LVKHSORRPPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSAKARG 120
Db 61 LVKHSORRPPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSAKARG 120

QY 121 DLGAFSRG 128
Db 121 DLGAFSRG 128

RESULT 3
US-10-687-361-16
; Sequence 16, Application US/10687361
; Publication No. US20040101896A1
; GENERAL INFORMATION:
; APPLICANT: Hunter, Tony
; APPLICANT: Lu, Kun Ping
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; FILE REFERENCE: 66671-078
; CURRENT APPLICATION NUMBER: US/10/687,361
; CURRENT FILING DATE: 2003-10-15
; PRIOR APPLICATION NUMBER: US 10/616,410
; PRIOR FILING DATE: 2003-07-08
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-687-361-16

Query Match      62.7%; Score 534; DB 16; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.4e-47;
Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 HLLVKHSORRPPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSAKA 118
Db 1 HLLVKHSORRPPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSAKA 60

QY 119 RGDLGAFSRGQMKPFEDASFALRTEGMSGPVFTDSGHIILRTE 163
Db 61 RGDLGAFSRGQMKPFEDASFALRTEGMSGPVFTDSGHIILRTE 105

RESULT 4
US-10-128-714-3443
; Sequence 3443, Application US/10128714
; Publication No. US20030119013A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroskin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 3443
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-3443

Query Match      51.3%; Score 437.5; DB 14; Length 178;
Best Local Similarity 51.5%; Pred. No. 3e-37;
Matches 88; Conservative 26; Mismatches 40; Indels 17; Gaps 3;

QY 7 LPGWKMSRSGRVYFNHITNASQWERPSG-----NSSSGK---NGOG 50
Db 9 LPAGWEVRHSNKNLPYFNPTKESRWEPSPSGTDELKLVYNAVHSAPAGRPDGTAG 68

QY 51 EPARYCRSHLLVKHSORRPPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQF 110
Db 69 E-GKIRCSEHLLKHDRSRPPSSWRREAITRSKEAEIILRGHEQIRSRGEVSLGDIASE 127

QY 111 SDCSSAKARGDLGAFSRGQMKPFEDASFALRTEGMSGPVFTDSGHIILR 161
Db 128 SDCSSARKKGDLGFFGRGEMQKFEFADAFALQPGVQSGIVETASGVHLLIER 178

RESULT 5
US-10-128-714-8443
; Sequence 8443, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wenqi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroskin, Alexey M
; APPLICANT: Lemieux, Sebastien M
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
```

```
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8443
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Aspergillus fumigatus
US-10-128-714-8443

Query Match          51.0%; Score 434.5; DB 14; Length 178;
Best Local Similarity 50.6%; Pred. No. 6.1e-37;
Matches 88; Conservative 26; Mismatches 43; Indels 17; Gaps 3;

QY 4 BEKLPWWEKMSRSGRVYFNHITNASQWRPSG-----NSSSGK--N 47
Db 6 ETGLPAGWEVRHSKNLPLYFNSTKESRWPESGDTLLKVMYMHYSAPAGPDGT 65
QY 48 GQGPAPVRCSHLLVVKHSQSRPPSSRQEKITRKKEALELINGYIQIKSGEEDFSLA 107
Db 66 AQGE-GKIRCSHLLIKHRDSRPPSSRWEAEITRSKEAEITLRGHEQIRSGEVSGLDIA 124
QY 108 SQFSDCSAKARGDLGAFSRQMKPPEADAFALRTGEMSGPVTDSGIHLIR 161
Db 125 VSESDCSSARKKDCGFFRGEMQKEFEDAAAFALQPGQGVIVETASGVHLIER 178

RESULT 6
US-10-263-828-116
; Sequence 116, Application US/10263828
; Publication No. US20030138905A1
; GENERAL INFORMATION:
; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Glenn, Matthew R.
; APPLICANT: Grigor, Murray R.
; APPLICANT: Molenaar, Adrian J.
; TITLE OF INVENTION: Compositions isolated from bovine
; TITLE OF INVENTION: mammary gland and methods for their use.
; FILE REFERENCE: 11000.1044UICON
; CURRENT APPLICATION NUMBER: US/10/263,828
; CURRENT FILING DATE: 2002-10-02
; NUMBER OF SEQ ID NOS: 136
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 116
; LENGTH: 87
; TYPE: PRT
; ORGANISM: Bovine
US-10-263-828-116

Query Match          47.7%; Score 406; DB 14; Length 87;
Best Local Similarity 96.1%; Pred. No. 2.2e-34;
Matches 74; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MADEKLPPEWKEKMSRSGRVYFNHITNASQWRPSGNSGGKNGGQGEFARVRCSHL 60
Db 1 MADEKLPPEWKEKMSRSGRVYFNHITNASQWRPSGNSGGKNGGQGEFARVRCSHL 60
QY 61 LVKHSQSRPPSSRQEK 77
Db 61 LVKHSQSRPPSSRQEK 77

RESULT 7
US-09-864-761-46074
; Sequence 46074, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
```

```
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46074
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008752.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.4
; OTHER INFORMATION: EST HUMAN HIT: BF314489.1, EVALUATE 2.00e-28
; OTHER INFORMATION: SWISSPROT HIT: Q13526, EVALUATE 2.00e-29
US-09-864-761-46074

Query Match          43.8%; Score 373; DB 9; Length 70;
Best Local Similarity 100.0%; Pred. No. 4.5e-31;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 RYYFNHITNASQWRPSGNSGGKNGGQGEFARVRCSHLLVVKHSQSRPPSSRQEKLTR 80
Db 1 RYYFNHITNASQWRPSGNSGGKNGGQGEFARVRCSHLLVVKHSQSRPPSSRQEKLTR 60
QY 81 TKEEALELIN 90
Db 61 TKEEALELIN 70

RESULT 8
US-10-342-555-2
; Sequence 2, Application US/10342555
; Publication No. US20030143615A1
; GENERAL INFORMATION:
; APPLICANT: HANES, Steven D.
```



```

; APPLICANT: DEVASAHAYAM, Gina
; APPLICANT: CHATURVEDI, Vishnu
; TITLE OF INVENTION: CAESI: A CANDIDA ALBICANS GENE, METHODS FOR MAKING AND
; TITLE OF INVENTION: USING, AND TARGETING IT OR ITS EXPRESSION PRODUCTS FOR
; TITLE OF INVENTION: ANTIFUNGAL APPLICATIONS
; FILE REFERENCE: 454311-2200.1
; CURRENT APPLICATION NUMBER: US/10/342,555
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: US/09/507,242
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 60/121,246
; PRIOR FILING DATE: 1999-02-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-342-555-2

Query Match 43.4%; Score 369.5; DB 14; Length 177;
Best Local Similarity 43.2%; Pred. No. 3.5e-30;
Matches 73; Conservative 34; Mismatches 49; Indels 13; Gaps 2;

QY 7 LPPGWEKMRSSRGVYFNFHITNASQWERPSGN-----SSSGKNGQGEPAR 54
Db 8 LPPNWTIRVSRSHNKKEYFLNQSTNESSWDPPYGTGKEVLNAYIAKFNKNGYKPLVNEGQ 67

QY 55 VRCSHLLVKHSQSRPPSWRQ-EKTRTKKEALELINGYIQIKSGEDFESLASQFSDC 113
Db 68 VRVSHLLIKNNQSRKPKSWKSPDGISRTDESQILKXHLERILSGEVLSELANTESDC 127

QY 114 SSARKARGDLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGHIILRT 162
Db 128 SSHDRGGDLGFFSKGQMPPEEAAFNLVHGVSNIIETNSGVHILQRT 176

RESULT 9
US-10-032-585-7374
; Sequence 7374, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7374
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7374

Query Match 43.4%; Score 369.5; DB 14; Length 177;
Best Local Similarity 43.2%; Pred. No. 3.5e-30;
Matches 73; Conservative 34; Mismatches 49; Indels 13; Gaps 2;

QY 7 LPPGWEKMRSSRGVYFNFHITNASQWERPSGN-----SSSGKNGQGEPAR 54
Db 8 LPPNWTIRVSRSHNKKEYFLNQSTNESSWDPPYGTGKEVLNAYIAKFNKNGYKPLVNEGQ 67

QY 55 VRCSHLLVKHSQSRPPSWRQ-EKTRTKKEALELINGYIQIKSGEDFESLASQFSDC 113
Db 68 VRVSHLLIKNNQSRKPKSWKSPDGISRTDESQILKXHLERILSGEVLSELANTESDC 127

QY 114 SSARKARGDLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGHIILRT 162
Db 128 SSHDRGGDLGFFSKGQMPPEEAAFNLVHGVSNIIETNSGVHILQRT 176
```

```

RESULT 10
US-10-320-797-3033
; Sequence 3033, Application US/10320797
; Publication No. US20040014955A1
; GENERAL INFORMATION:
; APPLICANT: Eroshkin, Alexey M.
; APPLICANT: Zamudio, Carlos
; TITLE OF INVENTION: IDENTIFICATION OF ESSENTIAL GENES OF CRYPTOCOCCUS NEOFORMANS AND
; TITLE OF INVENTION: METHODS OF USE
; FILE REFERENCE: 10182-021-999
; CURRENT APPLICATION NUMBER: US/10/320,797
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: 60/341,261
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 3361
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3033
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Cryptococcus neoformans
US-10-320-797-3033

Query Match 42.4%; Score 361; DB 15; Length 209;
Best Local Similarity 45.6%; Pred. No. 3.3e-29;
Matches 82; Conservative 23; Mismatches 43; Indels 32; Gaps 5;

QY 10 GWEKMRSSRGVYFNFHITNASQWERPSGNSS-----SSGKNGQ 49
Db 34 GWEIRFSNRQIFYPYNSERSISTWEPPELSAEQIQLPGAAYMNVQLAPAGGKEGQ 93

QY 50 GEPARVRCSHLLVKHSQSRPPSWR--QEKITRTKEALELIN---GYIQIKSGE--ED 102
Db 94 -----VRASHILAKHAGSRPPASWRNVDRKITTSDEAQAITEHQHAYLQSLPPADLPKE 148

QY 103 FESLASQFSDCSSAKARGDLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGHIILRT 162
Db 149 FAXIASTEDCSSARKGGDLGWFGRGQMKPFEDATFNTPVGQSLGIVKTDSGIHLILT 208

RESULT 11
US-10-424-599-185102
; Sequence 185102, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 185102
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_138161C.1.pap
US-10-424-599-185102

Query Match 37.3%; Score 318; DB 12; Length 126;
Best Local Similarity 52.8%; Pred. No. 5e-25;
Matches 67; Conservative 18; Mismatches 36; Indels 6; Gaps 2;

QY 41 SSSGKNGQGEPARVRCSHLLVKHSQSRPPSSNRQEK----ITRTKEALELINGYIQIK 96
Db 2 SSSSSKGGGG--TEVRASHILIKHEGSRKASWKPDEGRILKSTTRENAYVSQLALRDDI 59

QY 97 KSGEEDFESLASQFSDCSSAKARGDLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGI 156
```

Db 60 VSGKASFEDIASRFDSCSAKRGDLGPPGRGQMKPFEEATFALKVGEISDIVDTSKV 119
157 HIIRTE 163
120 HIIRTK 126

RESULT 12
US-10-424-599-185101
; Sequence 185101, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 185101
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_138160C.1.pap
US-10-424-599-185101

Query Match 36.9%; Score 314; DB 12; Length 126;
Best Local Similarity 52.0%; Pred. No. 1.3e-24;
Matches 66; Conservative 19; Mismatches 36; Indels 6; Gaps 2;

QY 41 SSSGKNGGGEAPRVRCSHLLVKHSQSRPPSSWRQK----ITRTKEALELINGYIQI 96
DB 2 SSSSKGGGG--LEVRASHILIKHEGSRKAKWKQDPGEIRIKSTTFENAVSQIKALRDI 59
QY 97 KSGEEDFSLAQSDCSAKARGDLGAFSRGQMKPFEDASFAIRTGMSGPFVTDGSI 156
DB 60 VSGKASFEDIASRFDSCSAKRGDLGPPGRGQMKPFEEATFALKIGISDIVDTSKV 119
QY 157 HIIRTE 163
DB 120 HIIRTK 126

RESULT 13
US-10-425-114-41877
; Sequence 41877, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 41877
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3136-013-F12_FLI.pap
US-10-425-114-41877

Query Match 36.0%; Score 306.5; DB 12; Length 188;
Best Local Similarity 46.5%; Pred. No. 1.3e-23;
Matches 66; Conservative 28; Mismatches 33; Indels 15; Gaps 4;
QY 35 ERPSGNSSSGK-----NGGGEPA-----RVRCSHLLVKHSQSRPPSSWRQK----ITR 80
DB 47 EHPS--SSSAGSRDRHHHHGRRPPSSADEKVRASHILIKHEGSRKAKSWRDPGVASAT 105
QY 81 TKEALELINGYIQIKSGEEDFSLAQSDCSAKARGDLGAFSRGQMKPFEDASFA 140
DB 106 TRDDAADLARALRQDIVSGDREFEDVAENSDCSAKRGDLGSGRGMQKPFKAAPA 165
QY 141 LRTGMSGPVTDSGIHILRT 162
DB 166 LKVGESIDVWDVTEGSHIIRKT 187

RESULT 14
US-10-425-114-65823
; Sequence 65823, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65823
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3354-123-A10_FLI.pap
US-10-425-114-65823

Query Match 36.0%; Score 306.5; DB 12; Length 204;
Best Local Similarity 46.5%; Pred. No. 1.5e-23;
Matches 66; Conservative 28; Mismatches 33; Indels 15; Gaps 4;
QY 35 ERPSGNSSSGK-----NGGGEPA-----RVRCSHLLVKHSQSRPPSSWRQK----ITR 80
DB 63 EHPS--SSSAGSRDRHHHHGRRPPSSADEKVRASHILIKHEGSRKAKSWRDPGVASAT 121
QY 81 TKEALELINGYIQIKSGEEDFSLAQSDCSAKARGDLGAFSRGQMKPFEDASFA 140
DB 122 TRDDAADLARALRQDIVSGDREFEDVAENSDCSAKRGDLGSGRGMQKPFKAAPA 181
QY 141 LRTGMSGPVTDSGIHILRT 162
DB 182 LKVGESIDVWDVTEGSHIIRKT 203

RESULT 15
US-10-687-361-17
; Sequence 17, Application US/10687361
; Publication No. US20040101896A1
; GENERAL INFORMATION:
; APPLICANT: Hunter, Tony
; APPLICANT: Lu, Kun Ping
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; FILE REFERENCE: 66671-078
; CURRENT APPLICATION NUMBER: US/10/687,361
; CURRENT FILING DATE: 2003-10-15
; PRIOR APPLICATION NUMBER: US 10/616,410
; PRIOR FILING DATE: 2003-07-08
; NUMBER OF SEQ ID NOS: 22

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2004, 14:35:24 ; Search time 259.925 seconds
(without alignments)
7747.025 Million cell updates/sec

Title: AAC50492

Perfect score: 474

Sequence: 1 ATGGCTGCTACGAGCAAC.....TTCGTGGAGAGCGTCTTAG 474

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseqn_25Jan04.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002s.*
- 7: Geneseqn2003as.*
- 8: Geneseqn2003bs.*
- 9: Geneseqn2003cs.*
- 10: Geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	474	100.0	474	3 AAC50492	Aac50492 Arabidops
2	338	71.3	507	3 AAC50504	Aac50504 Arabidops
3	318.6	67.2	531	3 AAC39615	Aac39615 Arabidops
4	301.2	63.5	621	3 AAC39654	Aac39654 Arabidops
5	213.6	45.1	459	3 AAC38179	Aac38179 Arabidops
6	208.6	44.0	411	3 AAC47703	Aac47703 Arabidops
7	208.6	44.0	643	3 AAC51199	Aac51199 Arabidops
8	208.6	44.0	674	3 AAC51182	Aac51182 Arabidops
9	208.6	44.0	677	3 AAC32694	Aac32694 Arabidops
10	207.6	43.8	1005	3 AAC47929	Aac47929 Arabidops
11	207.2	43.7	646	3 AAC33712	Aac33712 Arabidops
12	205.2	43.3	499	9 AD881685	Ad881685 Arabidops
13	200.8	42.4	699	6 ABK31041	Abk31041 Plant dwa
14	199	42.0	714	3 AAC51198	Aac51198 Arabidops
15	199	42.0	717	3 AAC34378	Aac34378 Arabidops
16	199	42.0	729	6 ABK31048	Abk31048 Plant dwa
17	196.6	41.5	404	7 ABX46988	Abx46988 Bovine ES
18	195.8	41.3	474	3 AAC41338	Aac41338 Zea mays
19	195.8	41.3	492	3 AAC35867	Aac35867 Zea mays
20	195.2	41.2	585	6 ABK31036	Abk31036 Plant dwa
21	194.2	41.0	447	3 AAC51749	Aac51749 Zea mays
22	193.4	40.8	487	3 AAC47648	Aac47648 Arabidops
23	193.4	40.8	669	6 ABK31042	Abk31042 Plant dwa

24	193.4	40.8	682	6 ABK30665	Abk30665 Plant dwa
25	193.4	40.8	682	6 ABL49479	Abi49479 Sequence
26	193.4	40.8	796	6 ABK30897	Abk30897 Plant dwa
27	192.4	40.6	972	3 AAC45109	Aac45109 Arabidops
28	192	40.5	360	6 ABQ85651	Abq85651 Arabidops
29	191.8	40.5	603	6 ABQ66368	Abq66368 Arabidops
30	191.8	40.5	640	3 AAC41085	Aac41085 Arabidops
31	191	40.3	450	6 ABK31037	Abk31037 Plant dwa
32	191	40.3	576	8 ACL22562	ACL22562 DNA clone
33	191	40.3	626	8 ACL22561	ACL22561 DNA clone
34	191	40.3	638	8 ACL22574	ACL22574 DNA clone
35	191	40.3	640	8 ACL22573	ACL22573 DNA clone
36	191	40.3	646	8 ACL22563	ACL22563 DNA clone
37	191	40.3	652	8 ACL22566	ACL22566 DNA clone
38	191	40.3	663	8 ACL22576	ACL22576 DNA clone
39	191	40.3	668	6 ABK31028	Abk31028 Plant dwa
40	191	40.3	672	8 ACL22565	ACL22565 DNA clone
41	191	40.3	678	8 ACL22564	ACL22564 DNA clone
42	191	40.3	688	8 ACL22559	ACL22559 DNA clone
43	191	40.3	693	8 ACL22560	ACL22560 DNA clone
44	191	40.3	712	8 ACL22570	ACL22570 DNA clone
45	190.4	40.2	668	3 AAC34045	Aac34045 Arabidops

ALIGNMENTS

RESULT 1

AAC50492

ID AAC50492 standard; DNA; 474 BP.

XX

AC AAC50492;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 65043.

XX

KW Hybridisation assay; Genetic mapping; Gene expression control;

KW protein identification; signal transduction pathway; metabolic pathway;

KW Promoter; termination sequence; ss.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR

PR 05-MAR-1999; 99US-0123180P.

PR

PR 09-MAR-1999; 99US-0123548P.

PR

PR 23-MAR-1999; 99US-0125788P.

PR

PR 25-MAR-1999; 99US-0126264P.

PR

PR 29-MAR-1999; 99US-0126785P.

PR

PR 01-APR-1999; 99US-0127462P.

PR

PR 06-APR-1999; 99US-0128234P.

PR

PR 08-APR-1999; 99US-0128714P.

PR

PR 16-APR-1999; 99US-0129845P.

PR

PR 19-APR-1999; 99US-0130077P.

PR

PR 21-APR-1999; 99US-0130449P.

PR

PR 23-APR-1999; 99US-0130510P.

PR

PR 28-APR-1999; 99US-0130891P.

PR

PR 30-APR-1999; 99US-0131449P.

PR

PR 30-APR-1999; 99US-0132048P.

PR

PR 04-MAY-1999; 99US-0132407P.

PR

PR 04-MAY-1999; 99US-0132484P.

PR

PR 05-MAY-1999; 99US-0132485P.

PR

PR 06-MAY-1999; 99US-0132486P.

PR

PR 06-MAY-1999; 99US-0132487P.

PR

PR 07-MAY-1999; 99US-0132863P.

PR

PR 11-MAY-1999; 99US-0134256P.

PR

PR 14-MAY-1999; 99US-0134218P.

PR

PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139422P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140911P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 10-AUG-1999; 99US-0147935P.
PR 11-AUG-1999; 99US-0148171P.
PR 13-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 22-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 28-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 28-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160800P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.


```
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 23-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147303P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-015753P.
PR 06-OCT-1999; 99US-0157865P.

PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 71.3%; Score 338; DB 3; Length 507;
Best Local Similarity 86.7%; Pred. No. 1.7e-97;
Matches 411; Conservative 0; Mismatches 0; Indels 63; Gaps 1;

Qy 1 ATGCTCGTACGAGCAAACTGCAAGAAATCACAAGGAAAGCTCCAGGAGAAAGCTCCAGGACTCTG 60
Db 56 ATGCTCGTACGAGCAAACTGCAAGAAATCACAAGGAAAGCTCCAGGAGAAAGCTCCAGGACTCTG 115
Qy 61 CTCGTACCAAGCGCGGAGGAAATCTGCGCCGACTACTGAGGAGTCAAGAAACCTCAC 120
Db 116 CTCGTACCAAGCGCGGAGGAAATCTGCGCCGACTACTGAGGAGTCAAGAAACCTCAC 175
Qy 121 CGTTACCGTCCCGAACCGTCGCTCTTCGTGAGATTCGTAATACCAAGAGACACAGAG 180
Db 176 CGTTACCGTCCCGAACCGTCGCTCTTCGTGAGATTCGTAATACCAAGAGACACAGAG 235
Qy 181 TTGTTGATCCGTAAACTTCCTTTTCAACGTCCTTGTTCGTAATCGCTCAAGATTACAAG 240
Db 236 TTGTTGATCCGTAAACTTCCTTTTCAACGTCCTTGTTCGTAATCGCTCAAGATTACAAG 294
Qy 241 ACGGATCTGAGATTCAGAGCCATGCGGTGTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
Db 295 ----- 294
Qy 301 TTGACGATCTGAGATTCAGAGCCATGCGGTGTAGCTCTTCAAGAGCTGCGGAGGCA 360
Db 295 --GACGATCTGAGATTCAGAGCCATGCGGTGTAGCTCTTCAAGAGCTGCGGAGGCA 352
Qy 361 TATTGTTGGTGTGTTGTTGAAGACACAAATCTCTGTGCCATTCATCAAGAGAGGTTACG 420
Db 353 TATTGTTGGTGTGTTGTTGAAGACACAAATCTCTGTGCCATTCATCAAGAGAGGTTACG 412
Qy 421 ATAATGCTAAAGATGTTCAATTTGCAAGAGGATTCGTTGAGAGCGGTGTTAG 474
Db 413 ATAATGCTAAAGATGTTCAATTTGCAAGAGGATTCGTTGAGAGCGGTGTTAG 466

RESULT 3
AAC39615
ID AAC39615 standard; DNA; 531 BP.
XX
```

[illegible]


```
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151085P.
PR 27-AUG-1999; 99US-0151086P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 06-OCT-1999; 99US-0157753P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 67.2%; Score 318.6; DB 3; Length 531;
Best Local Similarity 85.4%; Pred. No. 2.8e-91;
Matches 405; Conservative 0; Mismatches 5; Indels 64; Gaps 2;

QY 1 ATGGCTCGTACGAAGCAAACTGCAAGAAATCACACGGAGGAAGCTCCGAGGACTCTG 60
DB 89 ATGGCTCGTACGAAGCAAACTGCAAGAAATCACACGGAGGAAGCTCCGAGGACTCTG 148

QY 61 CTCGCTACCAAGGCGGAGGAATCTGCGCGACTACTGGAGGAGTCAAGAAACCTCAC 120
DB 149 CTCGCTACCAAGGCGGAGGAATCTGCGCGACTACTGGAGGAGTCAAGAAACCTCAC 208

QY 121 CGTTACCGTCCGGAACCGTCTGCTGAGTTGTAATACCAAGAGCACAGAG 180
DB 209 CGTTACCGTCCGGAACCGTCTGCTGAGTTGTAATACCAAGAGCACAGAG 268

QY 181 TTGTTGATCCGTAACCTTCCTTTTCAACGCTCTGTTGTTGTAATCGCTCAAGATTACAAG 240
DB 269 TTGTTGATCCGTAACCTTCCTTTTCAACGCTCTGTTGTTGTAATCGCTCAAGATTACAAG 328
```

```
QY 241 ACGGATCTGAGATTCAGAGCCATGCGGTCTTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
DB 329 ACGGATCTGAGATTCAGAGCCATGCGGTCTTAGCTCTTCAAGAGCTGCTGAA----- 382
QY 301 TTGACGATCTGAGATTCAGAGCCATGCGGTCTTAGCTCTTCAAGAGCTGCGGAGGCA 360
DB 383 -----GCA 385
QY 361 TATTGGTGGTTTGTGTAAGACACAAATCTCTGTGCCATTCATGCAAGAGGTTACG 420
DB 386 TATTGGTGGTTTGTGTAAGATACAAATCTCTGTGCCATTCATGCAAGAGGTTACG 445
QY 421 ATATGCTTAAAGATGTTCAATTTGCAAGAGGATTCGTGGAGGCGTCTTAG 474
DB 446 ATCATGCTTAAAGATGTTCAATTTGGC-AGAAGATTGCTGGAGGCGTCTTAG 498

RESULT 4
AAC39654
ID AAC39654 standard; DNA; 621 BP.
XX AAC39654;
XX 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 25421.
XX Arabidopsis thaliana.
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.
XX Arabidopsis thaliana.
XX EF1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-00301439.
XX 25-FEB-1999; 99US-0121825P.
XX 05-MAR-1999; 99US-0123180P.
XX 09-MAR-1999; 99US-0123548P.
XX 23-MAR-1999; 99US-0125788P.
XX 29-MAR-1999; 99US-0126264P.
XX 29-MAR-1999; 99US-0126785P.
XX 01-APR-1999; 99US-0127462P.
XX 08-APR-1999; 99US-0128234P.
XX 08-APR-1999; 99US-0128714P.
XX 16-APR-1999; 99US-0129845P.
XX 19-APR-1999; 99US-0130077P.
XX 21-APR-1999; 99US-0130449P.
XX 23-APR-1999; 99US-0130510P.
XX 23-APR-1999; 99US-0130891P.
XX 28-APR-1999; 99US-0131449P.
XX 30-APR-1999; 99US-0132048P.
XX 30-APR-1999; 99US-0132407P.
XX 04-MAY-1999; 99US-0132484P.
XX 05-MAY-1999; 99US-0132485P.
XX 06-MAY-1999; 99US-0132486P.
XX 06-MAY-1999; 99US-0132487P.
XX 07-MAY-1999; 99US-0132863P.
XX 11-MAY-1999; 99US-0134256P.
XX 14-MAY-1999; 99US-0134218P.
XX 14-MAY-1999; 99US-0134219P.
XX 14-MAY-1999; 99US-0134221P.
XX 14-MAY-1999; 99US-0134370P.
XX 18-MAY-1999; 99US-0134768P.
XX 19-MAY-1999; 99US-0134941P.
XX 20-MAY-1999; 99US-0135124P.
XX 21-MAY-1999; 99US-0135353P.
XX 24-MAY-1999; 99US-0135629P.
XX 25-MAY-1999; 99US-0136021P.
XX 27-MAY-1999; 99US-0136392P.
```

PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139453P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139811P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144082P.
PR 16-JUL-1999; 99US-0144083P.
PR 19-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0136782P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 09-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 23-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 25-AUG-1999; 99US-0149930P.
PR 26-AUG-1999; 99US-0150566P.
PR 27-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154033P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0156559P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 25-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 26-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 28-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

```
Query Match          63.5%; Score 301.2; DB 3; Length 621;
Best Local Similarity 81.9%; Pred. No. 1.1e-85;
Matches 388; Conservative 0; Mismatches 23; Indels 63; Gaps 1;

QY 1 ATGGCTCTGACGAAGCAAACTGCAAGAAAATACACGGAGGAAAGCTCCGAGGACTCTG 60
Db 92 ATGGCTCTGACGAAGCAAACTGCAAGAAAATACACGGAGGAAAGCTCCGAGGAGCAG 151
QY 61 CTCGCTACCAAGGGCGGAGGAAATCTGGCGGACTACTGAGGAGTCAAGAACTTCAAC 120
Db 152 CTCGCTACCAAGGGCGGAGGAAATCTGCAAGGACTACTGAGGAGTCAAGAACTTCAAC 211
QY 121 CGTTACCTCCCGGAACCGTCTGCTTTCGTCGAGATTCTGTAATACCAAGAGCACAGAG 180
Db 212 CGTTTCCTCTCGAAGCGTCTGCTCTTCGTGAGATTCTGTAATACCAAGAGCACAGAG 271
QY 181 TTGTTGATCGGTAACTTCTTTTCAACGCTTTGTTGCTGAAATGGCTCAAAGATTACAAG 240
Db 272 TTGTTGAACCGTAACTTCTTCCATTCACAGCTTTGTTGCTGAAATGGCTCAAAGATTCAA- 330
QY 241 ACGGATCTGAGATCCAGAGCCATCGGCTGTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
Db 331 ----- 330
QY 301 TTGACGGATCTGAGATCCAGAGCCATCGGCTGTAGCTCTTCAAGAGCTGCGGAGGCA 360
Db 331 --GACGGATCTGAGATCCAAAGCCATCGAGTGTAGCTCTTCAAGAGCTGCGGAGGCA 388
QY 361 TATTGCTGGGTTGTTTGAAGACACAAATCTGTGTCATTCATGCAAGAGGTTACG 420
Db 389 TATTGCTGGGTTGTTTGAAGACACAAATCTGTGTCATTCATGCAAGAGGTTACG 448
QY 421 ATAATGCTTAAAGATGTTCAATTTGCAAGAGGATTCGTGAGAGCTGCTGCTTAG 474
Db 449 ATTATGCTTAAAGATGTTCAATTTGCAAGAGGATTCGTGAGAGCTGCTGCTTAG 502

RESULT 5
AAC38179
ID AAC38179 standard; DNA; 459 BP.
AC AAC38179;
XX
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 20053.
XX
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX
PN EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
```

```
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134376P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140891P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 18-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
```

```
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145188P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145213P.
PR 27-JUL-1999; 99US-0145218P.
PR 27-JUL-1999; 99US-0145219P.
PR 28-JUL-1999; 99US-0145251P.
PR 02-AUG-1999; 99US-0145386P.
PR 02-AUG-1999; 99US-0145388P.
PR 02-AUG-1999; 99US-0145389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 23-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.

PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161408P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 45.1%; Score 213.6; DB 3; Length 459;
Best Local Similarity 84.3%; Pred. No. 9.6e-58;
Matches 252; Conservative 0; Mismatches 45; Indels 2; Gaps 1;

QY 1 ATGGCTCGTACGACCAAACTGCAAGAAATCACACGGAGGAAAGCTCCGAGGACTCTG 60
DB 97 ATGGCTCGTACGACCAAAACCGCTAGAAA--TCCACCGAGTAAAGCTCCGAGGAGCAG 154
QY 61 CTCGCTACCAAGCGCGGAGGAAATCTGGCCGACTACTGGAGGAGTCAAGAACTCTAC 120
DB 155 CTTGCCACCAAGCGCGGAGGAAATCAGCACCACACCGGAGGAGTCAAGAACTCTAC 214
QY 121 CGTTACCGTCCCGGAACCGTCTGCTTCTGAGATTCTGTAATATCAAGAGGACAGAG 180
DB 215 CGTTACCGTCCCGGAACCGTCTGCTTCTGAGATTCTGTAATATCAAGAGGAGTCTAA 274
QY 181 TTGTTGATCGGTAAACTCTTTTCAAGCTCTGTTGTTGTTGTTGTTGTTGTTGTTGTT 240
DB 275 TTGTTGATCCGCAAGCTTCCCTTCCAGGCTCTCGTTCGGGAAATAGCTCAGGACTTCAAG 334
QY 241 ACGGATCTGAGATTCAGAGCATCGCGTGTGTAGCTCTTCAAGAGCTGCTGAAGCATA 299
DB 335 ACGGATCTGAGGTTTCAGAGTCAAGGTCACCGGTTAGCTCTTCAAGAGCGGCTGAGGCATA 393

RESULT 6
AAC47703
ID AAC47703 standard; DNA; 411 BP.
XX
AC AAC47703;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 54805.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
```

PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127452P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136382P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139432P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139859P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140891P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 18-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-014819P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.

```

PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match      44.0%; Score 208.6; DB 3; Length 411;
Best Local Similarity 79.4%; Pred. No. 3.6e-56;
Matches 247; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1 ATGGCTGCTAGCAAGCAAACTGCAAGAAATACACGGAGGAAAGCTCCGAGGACTCTG 60
Db 1 ATGGCTGCTAGCAAGCAAACTGCAAGAAATACACGGAGGAAAGCTCCGAGGACTCTG 60
QY 61 CTCGTACCAAGGGCGGAGGAAATCTCGCGGACTACTGAGGAGTCAAGAACTTAC 120
Db 61 CTCGTACCAAGGGCGGAGGAAATCTCGCGGACTACTGAGGAGTCAAGAACTTAC 120
QY 121 CGTTACCGTCCCGGAACCGTCTTCTCGTGAGATTCCTAATACCAAGAGCACAG 180
Db 121 CGTTACCGTCCCGGAACCGTCTTCTCGTGAGATTCCTAATACCAAGAGCACAG 180
QY 181 TTGTTGATCCGTAACCTTCTTCAAGCTCTTGTGTAATCGCTCAAGATTACAAG 240
Db 181 TTGTTGATCCGTAACCTTCTTCAAGCTCTTGTGTAATCGCTCAAGATTACAAG 240
QY 241 ACGGATCTGAGATTCAGAGCATCGCTGTTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
Db 241 ACGGATCTGAGATTCAGAGCATCGCTGTTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
QY 301 TTGACGGATCT 311
Db 301 CTTGTTGGTCT 311

RESULT 7
AAC51199
ID AAC51199 standard; DNA; 643 BP.
AC AAC51199;
XX
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 67632.
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 67632.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway; metabolic pathway;
XX promoter; termination sequence; ss.

```

```

XX
OS Arabidopsis thaliana.
XX
XX EF1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 18-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.

```

```
PR 29-JUN-1999; 99US-0140981P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144326P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151085P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160982P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160982P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
Query Match 44.0%; Score 208.6; DB 3; Length 643;
Best Local Similarity 79.4%; Pred. No. 4.4e-56;
Matches 247; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 1 ATGGCTCGTACGAAGCAAACTGCAAGAAATCACACGAGGAAAGCTCCGAGGACTCTG 60
Db 91 ATGGCTCGTACGAAGCAAACTGCAAGAAATCACACGAGGAAAGCTCCGAGGACTCTG 150
QY 61 CTCGCTACCAAGCGCGGAGGAAATCTGCGCCGACTACTGGAGGAGTCAAGAACTCTAC 120
Db 151 CTTGCTACTAAGGCTCTGCTAAATCTGCACCAACTACTGTTGAGTCAAGAACTCTAG 210
QY 121 CGTTACGCTCCGAGACCGCTGCTCTTCGTGAGATTCGTAATACCAAGAGCAGAG 180
Db 211 CGTTACGCTCCGAGAACTGTTGCTCTCCGTGAATCCGTAAGTACCAAGAGTACTGAG 270
QY 181 TTGTTGATCCGTAACCTTCTTTTCAACGCTCTGTTGTAATCGCTCAAGATTACAAG 240
Db 271 TTGCTTATCAGGAAGTCTGCAATTCAGAGGCTAGTCCGCTGAGATTCCTCAAGATTCAAG 330
QY 241 ACGGATCTGAGATTCAGAGCCATCGGTGTTAGCTCTTTCAAGAGCTCTGAAGCATAT 300
Db 331 ACTGATTGCGTTTCCAGAGCCATGCTGCTTAGCTCTCCAGAGCTCTCAGAGCATAT 390
QY 301 TTGACCGATCT 311
Db 391 CTTGTTGGTCT 401
```



```
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149902P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151068P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
```

```
Query Match 44.0%; Score 208.6; DB 3; Length 674;
Best Local Similarity 79.4%; Pred. No. 4.5e-56;
Matches 247; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1 ATGGCTCGTACGAAGCAAACTGCAAGAAATCACAGGAGGAAAGCTCCGAGGACTCTG 60
DB 85 ATGGCTCGTACGAAGCAAACTGCTAAAGTCCACCGGAGGTAAAGCTCAAGGAGCA 144
QY 61 CTCGCTACCAAGCGCGGAGGAAATCTGCCGCACTACTGGAGGAGTCAAGAACTCTAC 120
DB 145 CTTCGTACTAAGCTCTCGTAAATCTGCACCACTACTGCTGAGTCAAGAACTCCAT 204
QY 121 CGTTACCGTCCGGAACCGCTCTCTCTGAGATTCGTAATACCAAGAGACACAG 180
DB 205 CGTTACCGTCTGGAACCTGTCTCTCCGTGAATCCGTAAGTACCAGAGAGTACTG 264
```

```
QY 181 TTGTGATCCGTAACCTTCTTCAACGCTTGTTCGTGAATCGCTCAAGATTCAAG 240
DB 265 TTGCTTATCAGAAACTCCATTTTCAGAGGCTAGTCGCTGAGATTGCTCAAGATTCAAG 324
QY 241 ACGGATCTGAGATTCAGAGCCATGCGGTGTAGCTCTTCAAGAAAGCTGCTGAAGCATAT 300
DB 325 ACTGATTGCGTTTCCAGAGCCATGCTGTCTTAGCTCTCCAGGAAGCTGCAAGAGCATAT 384
QY 301 TTGACGGATCT 311
DB 385 CTGTGTTGCT 395

RESULT 9
AAC32694
ID AAC32694 standard; DNA; 677 BP.
XX
AC AAC32694;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 305.
XX
KW Hybridisation assay; Genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
PF 25-FEB-2000; 2000EP-00301439.
PR 25-FEB-1999; 99US-0121825P.
PR 09-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
```

PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138884P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 17-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143422P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145132P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 25-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 07-OCT-1999; 99US-0157865P.
PR 08-OCT-1999; 99US-0158029P.
PR 12-OCT-1999; 99US-0158232P.
PR 13-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 44.0%; Score 208.6; DB 3; Length 677;
Best Local Similarity 79.4%; Pred. No. 4.5e-56;
Matches 247; Conservative 0; Mismatches 64; Indels 0; Gaps 0;


```
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149829P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.

PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 43.8%; Score 207.6; DB 3; Length 1005;
Best Local Similarity 79.1%; Pred. No. 1.1e-55;
Matches 246; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 ATGGCTCTGTAAGCAAACTGCAAGAAATCACCGAGGAAAGCTCCGAGGACTCTG 60
DB 373 ATGGCTCTGTAAGCAAACTGTAAGTCCACCGGAGGTAAGCTCAAGGAAGCAA 432
QY 61 CTCGCTACCAAGCGCGGAGGAAATCTGCGCGACTACTGGAGGAGTCAAGAACTCAC 120
DB 433 CTGTCTACTAAGCTGCTGTAATCTGCACCAACNACTGGTGGAGTCAAGAACTCAC 492
QY 121 CGTTACCGTCCCGAACCGTCGCTCTTCTGTGAGATTTCGTAATACCAAGAGCAGAG 180
DB 493 CGTTACCGTCCCGAACCTGTTGCTCTCCGTGAAATCCGTAAGTACCAAGAGTACTGAG 552
QY 181 TTGTTGATCCGTAAGAACTTCTTTTCAACGCTCTTTCGTGAAATCGCTCAAGATTACAAG 240
DB 553 TTGCTTATCAGGAACTGCCAATTCAGAGGCTAGTCCGTGAGATTGCTCAAGATTACAAG 612
QY 241 ACGGATCTGAGATTCCAGAGCCATCGGCTGTTAGCTCTTCAAGAGCTCTGAGCATAT 300
DB 613 ACTGATTGCTTTTCAGAGCCATGCTGTTAGCTCTTCCAGAGCTCGAAGCATAT 672
QY 301 TTGACGGATCT 311
DB 673 CTGTGTGCT 683

RESULT 11
AAC33712
ID AAC33712 standard; DNA; 646 BP.
XX
AC AAC33712;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 4049.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW Protein identification; signal transduction pathway; metabolic pathway;
KW Promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0121800P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
```

PR 23-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0143208P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.

PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149729P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-015659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-015753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159283P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.

```

PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 43.7%; Score 207; DB 3; Length 646;
Best Local Similarity 79.1%; Pred. No. 1.4e-55;
Matches 246; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 ATGGCTCGTACGAGCAAACTGCAAGAAAATCACGGAGGAAAGCTCCGAGGACTCTG 60
DB 94 ATGGCTCGTACGAGCAAACTGCAAGAAAATCACGGAGGAAAGCTCCGAGGAGCAA 153

QY 61 CTCGCTACCAAGCGCGAGGAATCTGCCGACTACTCGAGGACTCAAGAACTCTAC 120
DB 154 CTTGCTACTAAGGCTCTCGTAAATCTGCACCAACTACTGGTTGAGTCAAGAAACCAT 213

QY 121 CGTTACCGTCCGGAACCGTCTCTGTTGAGATTCGTAAATACCAAGAGACAGAG 180
DB 214 CGTTACCGTCTGGAATCTGCTCCGTTGAATCGTTAAGTACCAAGAGACTAG 273

QY 181 TTGTTGATCGTTAACTCTCTTTTCAAGCTCTGTTGCGTGAATCGTCAAGATTACAAG 240
DB 274 TTGCTTATCAGGAACCTGCCATTTCAGAGGCTAGTCCGTTGAGATTGCTCAAGATTCAAG 333

QY 241 ACGGATCTGAGATTCCAGAGCCATCGGTTGAGTCTTCAAGAGCTGCTGAAGCATAT 300
DB 334 ACTGATTGGTTTCAGAGCCATGCTGTTAGCTCTCCAGGAGCTGAGAGCATAT 393

QY 301 TTGACGGATCT 311
DB 394 CTTGTTGGTCT 404

RESULT 12
ID ADE81685
AC ADE81685 standard; cDNA; 499 BP.
AC ADE81685;
XX
XX
XX 29-JAN-2004 (first entry)
DE Arabidopsis thaliana expressed polynucleotide seq id 456.
KW genetically modified organism; transgenic organism; plant;
KW inhibitor testing; activator testing; modifier testing; fungicide;
KW insecticide; genetic function; genetic regulation; cellular metabolism;
KW gene; ss.
XX
XX Arabidopsis thaliana.
XX
XX US2003115639-A1.
PN
XX
XX 19-JUN-2003.
XX

```

```

PF 26-JAN-2001; 2001US-00770961.
XX
PR 27-JAN-2000; 2000US-0178466P.
XX
XX
PA (GORL/) GORLACH J.
PA (ANY/) AN Y.
PA (HAMI/) HAMILTON C M.
PA (PRIC/) PRICE J L.
PA (RAIN/) RAINES T W.
PA (YUY/) YU Y.
PA (RAME/) RAMEAKA J G.
PA (PAGE/) PAGE A.
PA (MATH/) MATHAW A V.
PA (LEDF/) LEDFORD B L.
PA (WOES/) WOESSNER J P.
PA (HAAS/) HAAS W D.
PA (GARC/) GARCIA C A.
PA (KRIC/) KRICKER M.
PA (SLAT/) SLATER T.
PA (DAVI/) DAVIS K R.
PA (ALLE/) ALLEN K.
PA (HOFF/) HOFFMAN N.
PA (HURB/) HURBAN P.

XX
XX Goriach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
PI Hurban P;
XX
XX WPI; 2003-810930/76.
XX
PT Novel Arabidopsis thaliana nucleic acids useful for generating
PT genetically modified transgenic organisms, for screening biologically
PT active agents such as fungicides, insecticides.
XX
XX Claim 1; SEQ ID NO 456; 44pp; English.
XX
XX The invention describes a nucleic acid (1) comprising a sequence capable
XX of hybridising under stringent conditions to any one of 99 fully defined
XX Arabidopsis thaliana sequences (1) as given in specification e.g., 360,
XX 1137, 455, 219, 472 nucleotides, etc, or its fragment. (1) is useful as a
XX hybridisation probe to complementary molecules in a cDNA library. (1) is
XX also useful for generating genetically modified and transgenic organisms,
XX usually plant cells and plants. A protein encoded by (1) is useful in
XX screening assays to determine the effect of candidate inhibitors,
XX activators or modifiers of the gene product. The protein is also useful
XX for screening biologically active agents e.g., fungicides and
XX insecticides. A genetically modified cell, comprising an exogenous
XX nucleic acid, where the nucleic acid comprises transcription regulatory
XX sequences operably linked to a sequence capable of hybridising under
XX stringent conditions to (1) is useful in the study of genetic function
XX and regulation, for alteration of the cellular metabolism and for
XX screening compounds that may affect the biological function of the gene
XX or gene product. This sequence represents an Arabidopsis thaliana
XX polynucleotide of the invention.
XX
XX Sequence 499 BP; 146 A; 121 C; 111 G; 116 T; 0 U; 5 Other;

Query Match 43.3%; Score 205.2; DB 9; Length 499;
Best Local Similarity 78.4%; Pred. No. 4.8e-55;
Matches 243; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 2 TGCGCTCGTACGAGCAAACTGCAAGAAAATCACCGAGGAAAGCTCCGAGGACTCTGC 61
DB 80 TGNNTGNCCAGCAAGCAACCGCTCGTAACTCCACCGAGGTAAGCTCCAGGAAGCAAC 139
QY 62 TCCTACCAAGCGCGAGGAAATCTGCCGACTACTCGAGGAGTCAAGAAACCTCAAC 121
DB 140 TTGCTACTAAGGCTGCTCGTAAATCTGCACCAACTACTCGTGGAGTCAAGAAACCATC 199
QY 122 GTTACCGTCCCGAACCGTCGCTCTTCGTGAGTTCTGTAATACCAAGAGACAGAGT 181
DB 200 GTTACCGTCCGAACTGTTGCTCTCCGTGAAATCCGTAAGTACCAAGAGAGTACTGAGT 259

```

QY 182 TGTGATCCGTAACCTTCTTCAACGCTTGTTCGTAATCGCTCAAGATTACAGA 241
 |||||
 Db 260 TGCATTATCAGGAACGCAATTCAGAGGCTAGTTCTGAGATTCTCAAGATTCAAGA 319
 |||||
 QY 242 CGGATCTCAGATTCCAGAGCCATGCGGTGTAGCTCTTCAAGAAGCTCTGAAGCATATT 301
 |||||
 Db 320 CTGATTTCGGTTTCCAGAGCCATGCTGTAGCTCTCAGGAAGCTCGAGAGCATATC 379
 |||||
 QY 302 TGAAGGATCT 311
 |||||
 Db 380 TTGTTGGTCT 389
 |||||

RESULT 13
 ID ABK31041
 AC ABK31041 standard; cDNA; 699 BP.
 XX ABK31041;
 XX 07-AUG-2003 (revised)
 DT 23-APR-2002 (first entry)
 XX
 DE Plant dwarfing/stunting related cDNA seq ID 455.
 XX
 KW Plant; ss; dwarfism; stunting; EST; expressed sequence tag;
 KW transgenic plant; plant metabolism.
 OS Unidentified.
 XX
 PN WO200208410-A2.
 XX
 PD 31-JAN-2002.
 XX
 PF 20-JUL-2001; 2001WO-US023120.
 XX
 PR 20-JUL-2000; 2000US-0219809P.
 PR 20-JUL-2000; 2000US-0219810P.
 XX
 PA (DOWC) DOW CHEM CO.
 PA (REDDY) REDDY S. A.
 PA (LARR) LARRINDA M I.
 PA (RUEG) RUEGGER M.
 PA (WGL) WEGELARZ T.
 PA (BLAK) BLAKESLEE B.
 PA (ORIE) ORLEDO V B J.
 PA (SAVI) SAVICKAS J P.
 PA (MCCR) MCCREY A D.
 PA (MILL) MILLER A B.
 PA (POGU) POGUE P G.
 PA (DELL) DELLA-CIOPPA R G.
 PA (WOLF) WOLFE M G.
 PA (ZHEN) ZHENG W.
 PA (GACH) GACHOTTE D.
 PA (GROS) GROSLEY R.
 PA (PELL) PELL R.
 XX
 PI Reddy SA, Larrinua MI, Ruegger M, Weglarz T, Blakeslee B;
 PI Oriedo VBJ, Savickas JP, McCreery AD, Miller AB, Pogue PG;
 PI Della-Cioppa RG, Wolfe MG, Zheng W, Gachotte D, Grosley R, Pell R;
 XX
 DR WPI; 2002-164823/21.
 XX
 PT Polynucleotide and amino acid sequences identified in one or more
 PT metabolic pathways that lead to dwarfism and stunting in plants, useful
 PT in agriculture to create dwarf varieties of any plant species.
 XX
 PS Disclosure; Fig 9; 717pp; English.
 XX
 CC The invention relates to polynucleotide and amino acid sequences
 CC identified in one or more metabolic pathways that lead to dwarfism and
 CC stunting in plants. Also included are vectors comprising the
 CC polynucleotides, transgenic plants (including the seed and leaf)

CC transfected with the polynucleotides or vectors, a process for altering
 CC the metabolism of a plant comprising providing the above vector and a
 CC plant, and transfecting the plant with the vector under conditions such
 CC that the metabolism of the plant is altered by expression of the isolated
 CC nucleic acid from the vector, e.g. such that a stunting phenotype in an
 CC industrial plant is produced and a process for the characterisation of
 CC fractionated biological samples, comprising (a) providing one or more
 CC fractionated biological samples, references samples, a gas chromatography
 CC apparatus, a mass spectroscopy apparatus or data analysis software and
 CC (b) treating the fractionated biological samples and the reference
 CC samples with the gas chromatography apparatus to generate chromatographic
 CC data corresponding to the fractionated biological samples and the
 CC reference samples, (c) treating the fractionated biological samples and
 CC the reference samples with the mass spectroscopy apparatus to generate
 CC spectroscopic data corresponding to the fractionated biological samples
 CC and the reference samples and (d) processing the chromatographic and the
 CC spectroscopic data with the data analysis software. The nucleic acid and
 CC the vector are useful for altering the metabolism of a plant and for
 CC stunting a plant. The nucleic acids are useful in agriculture to create
 CC dwarf varieties of any plant species. The present sequence is a plant
 CC cDNA contig or singleton (related to dwarfism/stunting) identified by
 CC searching a nucleic acid database with plant EST (expressed sequence tag)
 CC and a BLAST (basic local alignment tool) stringency of e-20. (Updated on
 CC 07-AUG-2003 to correct OS field.)
 XX
 SQ Sequence 699 BP; 189 A; 130 C; 178 G; 201 T; 0 U; 1 Other;
 Query Match 42.4%; Score 200.8; DB 6; Length 699;
 Best Local Similarity 78.0%; Pred. No. 1.4e-53;
 Matches 241; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 3 GGCTCGTAGAGCAAACTGCAAGAAATCACACGGAGGAAAGCTCGAGGACTCTGCT 62
 Db 64 GNTCTGTACCAAGCAAACTGCTCGTAAGTCTACCGAGGAAAGGACCTAGGAAGCAACT 123
 QY 63 CGCTACCAAGGCGGAGGAAATCTGCGCCGACTACTGGAGGAGTCAAGAAACCTCACCG 122
 Db 124 TGCTACTAAGGCTGCTGTAAGTCTGCTCTCTACTACTTGGTGGAGTAAAGAACCTCACAG 183
 QY 123 TTACGTCCTCCGAACCGTCGCTCTTCGTGAGATTCGTAAATACCAGAGAGCAGAGTT 182
 Db 184 ATACCGCCCTGTGTACTGTTGCTCTTCGTGAAATCCGTAAGTACCAGAGAGTACTGAGCT 243
 QY 183 GTTGATCCGTAAACTTCCTTTTCAACGCTTGTGTTGCTCAAGATCGCTCAAGATTACAAGAC 242
 Db 244 CTTGATCAGGAAGCTCCCATTCAGAGGCTTGTTCGTGAANTTCTCAGGATTTCAAGAC 303
 QY 243 GGATCTGAGATTCAGAGCCATGCGGTGTTAGCTCTTCAAGAGCTGCTGAAGCATATT 302
 Db 304 TGATCTGGTTCAGAGATCATGCTGTTGGCTCTTCAGGAGGCTGCTGAGGCATACTT 363
 QY 303 GACGGATCT 311
 Db 364 GGTGGTCT 372
 |||||

RESULT 14
 AAC51198
 ID AAC51198 standard; DNA; 714 BP.
 XX
 AC AAC51198;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 67629.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway; metabolic pathway;
 KW promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.

PR	22-SEP-1999;	99US-0155139P.	AC	AAC34378;
PR	23-SEP-1999;	99US-0155486P.	XX	
PR	24-SEP-1999;	99US-0155659P.	DT	17-OCT-2000 (first entry)
PR	28-SEP-1999;	99US-0156458P.	XX	
PR	29-SEP-1999;	99US-0156596P.	DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 6434.
PR	04-OCT-1999;	99US-0157117P.	XX	
PR	05-OCT-1999;	99US-0157753P.	KW	Hybridisation assay; Genetic mapping; Gene expression control;
PR	06-OCT-1999;	99US-0157865P.	KW	protein identification; signal transduction pathway; metabolic pathway;
PR	07-OCT-1999;	99US-0158029P.	KW	promoter; termination sequence; ss.
PR	08-OCT-1999;	99US-0158232P.	XX	
PR	12-OCT-1999;	99US-0158369P.	OS	Arabidopsis thaliana.
PR	13-OCT-1999;	99US-0159293P.	XX	
PR	13-OCT-1999;	99US-0159294P.	PN	EP1033405-A2.
PR	13-OCT-1999;	99US-0159295P.	XX	
PR	14-OCT-1999;	99US-0159329P.	PD	06-SEP-2000.
PR	14-OCT-1999;	99US-0159330P.	XX	
PR	14-OCT-1999;	99US-0159331P.	PF	25-FEB-2000; 2000EP-00301439.
PR	14-OCT-1999;	99US-0159637P.	XX	
PR	14-OCT-1999;	99US-0159638P.	PR	25-FEB-1999; 99US-0121825P.
PR	18-OCT-1999;	99US-0159584P.	PR	05-MAR-1999; 99US-0123180P.
PR	21-OCT-1999;	99US-0160741P.	PR	09-MAR-1999; 99US-0123548P.
PR	21-OCT-1999;	99US-0160767P.	PR	23-MAR-1999; 99US-0125788P.
PR	21-OCT-1999;	99US-0160768P.	PR	25-MAR-1999; 99US-0126264P.
PR	21-OCT-1999;	99US-0160768P.	PR	29-MAR-1999; 99US-0126785P.
PR	21-OCT-1999;	99US-0160770P.	PR	01-APR-1999; 99US-0127462P.
PR	21-OCT-1999;	99US-0160815P.	PR	06-APR-1999; 99US-0128234P.
PR	21-OCT-1999;	99US-0160814P.	PR	08-APR-1999; 99US-0128714P.
PR	22-OCT-1999;	99US-0160980P.	PR	16-APR-1999; 99US-0129845P.
PR	22-OCT-1999;	99US-0160981P.	PR	19-APR-1999; 99US-0130077P.
PR	22-OCT-1999;	99US-0160989P.	PR	21-APR-1999; 99US-0130449P.
PR	25-OCT-1999;	99US-0161404P.	PR	23-APR-1999; 99US-0130510P.
PR	25-OCT-1999;	99US-0161405P.	PR	23-APR-1999; 99US-0130891P.
PR	25-OCT-1999;	99US-0161406P.	PR	28-APR-1999; 99US-0131449P.
PR	26-OCT-1999;	99US-0161359P.	PR	30-APR-1999; 99US-0132048P.
PR	26-OCT-1999;	99US-0161360P.	PR	30-APR-1999; 99US-0132407P.
PR	26-OCT-1999;	99US-0161361P.	PR	04-MAY-1999; 99US-0132484P.
PR	28-OCT-1999;	99US-0161920P.	PR	05-MAY-1999; 99US-0132485P.
PR	28-OCT-1999;	99US-0161992P.	PR	06-MAY-1999; 99US-0132486P.
PR	28-OCT-1999;	99US-0161993P.	PR	06-MAY-1999; 99US-0132487P.
PR	29-OCT-1999;	99US-0162142P.	PR	07-MAY-1999; 99US-0132863P.
Query Match 42.0%; Score 199; DB 3; Length 714;			PR	11-MAY-1999; 99US-0134256P.
Best Local Similarity 77.5%; Pred. No. 5.4e-53;			PR	14-MAY-1999; 99US-0134218P.
Matches 241; Conservative 0; Mismatches 70; Indels 0; Gaps 0;			PR	14-MAY-1999; 99US-0134219P.
QY	1	ATGGCTCGTACCAAGCAAACTGCAGAAATCACACGAGGAAAGCTCCGAGGACTCTG 60	PR	14-MAY-1999; 99US-0134221P.
Db	64	ATGGCTCGTACTAGCAACACCTGTAAGTTACTGGAGGAAAGGCTCCTAGGAGCAG 123	PR	14-MAY-1999; 99US-0134370P.
QY	61	CTCGCTACCAAGCGGCGAGGAAATCTGCCGACTTCTGGAGGAGTCAAGAAACCTCAC 120	PR	18-MAY-1999; 99US-0134768P.
Db	124	CTTGCTACAAAGGCTGCAGTAAGTCTGCACCAACCACTGGAGGAGTCAAGAAAGCCCAT 183	PR	19-MAY-1999; 99US-0134941P.
QY	121	CGTTACCGTCCGGRACCGCTCTTCTGTGAGATTCGTAATACCAGAGACACAGAG 180	PR	20-MAY-1999; 99US-0135124P.
Db	184	CGTTACCGTCCGGRACCGCTCTTCTGTGAGATTCGTAATACCAGAGAGTACCGAG 243	PR	21-MAY-1999; 99US-0135353P.
QY	181	TTGTGATCCGTAACCTCTCTTTTCAACGCTTTGTTGCTGAAATCGCTCAAGATTACAAG 240	PR	24-MAY-1999; 99US-0135629P.
Db	244	TTGTGATCAGGAGGCTCCCTTTCCAGAGGCTAGTTCTGTGAGATTGCCAGATTTCAG 303	PR	25-MAY-1999; 99US-0136021P.
QY	241	ACGGATCTGAGATTCAGAGCCATCGGTTGATGCTTTCAAGAGCTCTGTAAGCATAT 300	PR	27-MAY-1999; 99US-0136392P.
Db	304	ACTGACTTGGTTTCCAGAGCCATGCTGTGCTTGCTGCACTCCAGGAGGCTCTGAAGCATAC 363	PR	28-MAY-1999; 99US-0136782P.
QY	301	TTGACCGATCT 311	PR	01-JUN-1999; 99US-0137222P.
Db	364	CTGTGGGTCT 374	PR	03-JUN-1999; 99US-0137528P.
RESULT 15			PR	04-JUN-1999; 99US-0137502P.
AAC34378			PR	07-JUN-1999; 99US-0137724P.
ID AAC34378 standard; DNA; 717 BP.			PR	08-JUN-1999; 99US-0138094P.
XX			PR	10-JUN-1999; 99US-0138847P.
			PR	14-JUN-1999; 99US-0139119P.
			PR	16-JUN-1999; 99US-0139452P.
			PR	17-JUN-1999; 99US-0139453P.
			PR	18-JUN-1999; 99US-0139454P.
			PR	18-JUN-1999; 99US-0139455P.
			PR	18-JUN-1999; 99US-0139456P.
			PR	18-JUN-1999; 99US-0139457P.
			PR	18-JUN-1999; 99US-0139458P.
			PR	18-JUN-1999; 99US-0139459P.
			PR	18-JUN-1999; 99US-0139460P.
			PR	18-JUN-1999; 99US-0139461P.
			PR	18-JUN-1999; 99US-0139462P.

Tue Jun 29 08:37:07 2004

PR	18-JUN-1999;	99US-0139463P;
PR	18-JUN-1999;	99US-0139750P;
PR	18-JUN-1999;	99US-0139763P;
PR	21-JUN-1999;	99US-0139817P;
PR	22-JUN-1999;	99US-0139899P;
PR	22-JUN-1999;	99US-0140353P;
PR	23-JUN-1999;	99US-0140354P;
PR	23-JUN-1999;	99US-0140695P;
PR	24-JUN-1999;	99US-0140823P;
PR	28-JUN-1999;	99US-0140991P;
PR	29-JUN-1999;	99US-0141287P;
PR	30-JUN-1999;	99US-0141842P;
PR	01-JUL-1999;	99US-0142154P;
PR	02-JUL-1999;	99US-0142055P;
PR	06-JUL-1999;	99US-0142803P;
PR	08-JUL-1999;	99US-0142320P;
PR	09-JUL-1999;	99US-0142977P;
PR	12-JUL-1999;	99US-0143542P;
PR	13-JUL-1999;	99US-0143624P;
PR	14-JUL-1999;	99US-0144005P;
PR	15-JUL-1999;	99US-0144086P;
PR	16-JUL-1999;	99US-0144325P;
PR	19-JUL-1999;	99US-0144331P;
PR	19-JUL-1999;	99US-0144332P;
PR	19-JUL-1999;	99US-0144333P;
PR	19-JUL-1999;	99US-0144334P;
PR	19-JUL-1999;	99US-0144335P;
PR	20-JUL-1999;	99US-0144352P;
PR	20-JUL-1999;	99US-0144632P;
PR	20-JUL-1999;	99US-0144884P;
PR	21-JUL-1999;	99US-0144814P;
PR	21-JUL-1999;	99US-0145086P;
PR	21-JUL-1999;	99US-0145088P;
PR	22-JUL-1999;	99US-0145087P;
PR	22-JUL-1999;	99US-0145085P;
PR	22-JUL-1999;	99US-0145089P;
PR	22-JUL-1999;	99US-0145192P;
PR	23-JUL-1999;	99US-0145145P;
PR	23-JUL-1999;	99US-0145218P;
PR	23-JUL-1999;	99US-0145224P;
PR	26-JUL-1999;	99US-0145276P;
PR	27-JUL-1999;	99US-0145913P;
PR	27-JUL-1999;	99US-0145918P;
PR	27-JUL-1999;	99US-0145919P;
PR	28-JUL-1999;	99US-0145951P;
PR	02-AUG-1999;	99US-0146386P;
PR	02-AUG-1999;	99US-0146388P;
PR	02-AUG-1999;	99US-0146389P;
PR	03-AUG-1999;	99US-0147038P;
PR	04-AUG-1999;	99US-0147204P;
PR	09-AUG-1999;	99US-0147302P;
PR	05-AUG-1999;	99US-0147192P;
PR	06-AUG-1999;	99US-0147260P;
PR	06-AUG-1999;	99US-0147303P;
PR	06-AUG-1999;	99US-0147416P;
PR	09-AUG-1999;	99US-0147493P;
PR	09-AUG-1999;	99US-0147933P;
PR	10-AUG-1999;	99US-0148171P;
PR	11-AUG-1999;	99US-0148319P;
PR	12-AUG-1999;	99US-0148341P;
PR	13-AUG-1999;	99US-0148565P;
PR	13-AUG-1999;	99US-0148684P;
PR	16-AUG-1999;	99US-0149368P;
PR	17-AUG-1999;	99US-0149175P;
PR	18-AUG-1999;	99US-0149426P;
PR	20-AUG-1999;	99US-0149722P;
PR	20-AUG-1999;	99US-0149723P;
PR	20-AUG-1999;	99US-0149929P;
PR	23-AUG-1999;	99US-0149902P;
PR	23-AUG-1999;	99US-0149930P;
PR	25-AUG-1999;	99US-0150566P;

26-AUG-1999;	99US-0150894P.
27-AUG-1999;	99US-0151065P.
27-AUG-1999;	99US-0151066P.
27-AUG-1999;	99US-0151080P.
30-AUG-1999;	99US-0151303P.
30-AUG-1999;	99US-0151438P.
31-SEP-1999;	99US-0151930P.
07-SEP-1999;	99US-0152363P.
10-SEP-1999;	99US-0153070P.
13-SEP-1999;	99US-0153758P.
15-SEP-1999;	99US-0154018P.
16-SEP-1999;	99US-0154039P.
20-SEP-1999;	99US-0154779P.
22-SEP-1999;	99US-0155139P.
23-SEP-1999;	99US-0155486P.
23-SEP-1999;	99US-0155659P.
28-SEP-1999;	99US-0156458P.
29-SEP-1999;	99US-0156596P.
04-OCT-1999;	99US-0157117P.
05-OCT-1999;	99US-0157753P.
06-OCT-1999;	99US-0157865P.
07-OCT-1999;	99US-0158029P.
08-OCT-1999;	99US-0158232P.
12-OCT-1999;	99US-0158363P.
13-OCT-1999;	99US-0159293P.
13-OCT-1999;	99US-0159294P.
13-OCT-1999;	99US-0159295P.
14-OCT-1999;	99US-0159328P.
14-OCT-1999;	99US-0159330P.
14-OCT-1999;	99US-0159331P.
14-OCT-1999;	99US-0159637P.
14-OCT-1999;	99US-0159638P.
18-OCT-1999;	99US-0159584P.
21-OCT-1999;	99US-0160741P.
21-OCT-1999;	99US-0160747P.
21-OCT-1999;	99US-0160768P.
21-OCT-1999;	99US-0160770P.
21-OCT-1999;	99US-0160814P.
21-OCT-1999;	99US-0160815P.
22-OCT-1999;	99US-0160980P.
22-OCT-1999;	99US-0160981P.
22-OCT-1999;	99US-0160982P.
22-OCT-1999;	99US-0161404P.
25-OCT-1999;	99US-0161405P.
25-OCT-1999;	99US-0161406P.
26-OCT-1999;	99US-0161359P.
26-OCT-1999;	99US-0161360P.
26-OCT-1999;	99US-0161361P.
28-OCT-1999;	99US-0161320P.
28-OCT-1999;	99US-0161992P.
28-OCT-1999;	99US-0161993P.
28-OCT-1999;	99US-0162142P.

Query Match 42.0%; Score 199; DB 3; Length 717;

Best Local Similarity 77.5%; Pred. No. 5.4e-53;

Best local similarity					
Matches	241;	Conservative	0;	Mismatches	70;
					Indels
					Gaps
					0;

CONCLUSIONS

1 ATGGCTCGTACGAAGCAAACTGCAAGAAATCACACGGAGGAAAGCTCCGAGGACTCTTG 60

1. The first part of the document is a title page. It contains the title "THE HISTORY OF THE UNITED STATES OF AMERICA" and the author "BY JAMES MADISON".

b
66 ATGGCTCGTACTAAGCAAACAGCTCGTAAGTCTACTGGAGGAAGGCTCCTAGGAAGCAG 125

61 CTCGCTACCAAGCGCGGAGGAAATCTGCGCGGACTACTGGAGGAGTCAAGAAACCTCAC 120

198 E

b
126 CTTGCTACAAAGGCTGCACGTAAGTCTGCACCAACCACTGGAGGAGTCAAGAAAGCCCCAT 183

[illegible]

by
121 CGTTACCGTCCCGGAACCGTCGCTCTTGAGATTCGTAATAACCAAAAGACACAAG 138

1-2
ACCTCCGCGCAGTTCCTCAATTCTGTGGTAACAGACAGTACCCAG

245

186 GGTACCGTCCAGGAACGTTGGCACACGTCGTAATCGTAACTACCACTAAGCATTGCC

181 AACCTCATCCCTTAAAGTCCTGTTGGTGAAATCGCTCAAGATTACAAG 240

[illegible]

246 TTCCTGATCAGACGCTCCCTTCCAGAGGCTAGTTCGTGAGATTGCCAGGATTCAAG 305

[illegible]

Tue Jun 29 08:37:07 2004

QY	241	ACGGATCTGAGATCCAGAGCCATGCGGTGTTAGCTCTTCAAGAGCTGCTGAGCATAT	300
Db	306	ACTGACTTGGCGTTCCAGAGCCATGCTGTGCTTGCACTCCAGGAGGCTGCTGAGCATAC	365
QY	301	TTGACGGATCT	311
Db	366	CTGTGGGTCT	376

Search completed: June 28, 2004, 16:36:37
Job time : 263.425 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 16:23:14 ; Search time 129 Seconds
(without alignments)
357.017 Million cell updates/sec

Title: AAC50492

Perfect score: 852

Sequence: 1 MADEKLPQWEXKMSRSSG.....GEMSGPVFTDSGHIILRTTE 163

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseq1980s:*
2: Geneseq1990s:*
3: Geneseq2000s:*
4: Geneseq2001s:*
5: Geneseq2002s:*
6: Geneseq2003as:*
7: Geneseq2003bs:*
8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	852	100.0	163	2 AAW18312	Aw18312 NIMA-inte
2	852	100.0	163	4 AAG66532	Aag66532 Human int
3	852	100.0	163	7 ADC72966	Adc72966 Human pet
4	734	86.2	163	2 AAY48377	Aay48377 Human pro
5	705	82.7	131	4 ABG12572	Abg12572 Novel hum
6	642	75.4	259	4 ABG11947	Abg11947 Novel hum
7	580	68.1	195	4 AAU32052	Aau32052 Novel hum
8	456.5	53.6	166	4 ABB66150	Abb66150 Drosophil
9	437.5	51.3	178	6 ABJ25785	Abj25785 Aspergill
10	434.5	51.0	178	6 ABJ26385	Abj26385 Aspergill
11	406	47.7	87	4 AAB87657	Aab87657 Bovine ma
12	373	43.8	70	4 ABB43821	Abb43821 Peptide #
13	373	43.8	70	4 AAM37729	Aam37729 Peptide #
14	373	43.8	70	4 AAM77544	Aam77544 Human bon
15	373	43.8	70	4 AAM64786	Aam64786 Human bra
16	373	43.8	70	5 ABG46568	Abg46568 Human pep
17	369.5	43.4	177	3 AAY95876	Aay95876 Candida a
18	369.5	43.4	177	5 ABP73537	Abp73537 Candida a
19	361	42.4	209	7 ADB69989	Adb69989 C. neofo
20	301	35.3	119	3 AAG43433	Aag43433 Arabidops
21	301	35.3	119	3 AAG06627	Aag06627 Arabidops
22	301	35.3	119	3 AAG14662	Aag14662 Arabidops
23	233	27.3	35	5 ABP35511	Abp35511 Human iso
24	215	25.2	59	3 AAB21975	Aab21975 Pin1/huma
25	191	22.4	34	4 AAB74938	Aab74938 Peptidyl

26	186	21.8	155	4 ABG11946	Abg11946 Novel hum
27	179.5	21.1	92	3 AAY90945	Aay90945 Cenarchae
28	179	21.0	114	4 ABG12571	Abg12571 Novel hum
29	172	20.2	31	3 AAB21943	Aab21943 Pin1/huma
30	158	18.5	142	3 AAG11000	Aag11000 Arabidops
31	158	18.5	169	3 AAG10999	Aag10999 Arabidops
32	157.5	18.5	130	4 ABB68428	Abb68428 Drosophil
33	152	17.8	131	2 AAW79203	Aaw79203 Human par
34	152	17.8	131	5 ABB05432	Abb05432 Human par
35	143	16.8	138	3 AAG00864	Aag00864 Human sec
36	138	16.2	92	5 AAO18579	Aao18579 E coli pa
37	137	16.1	333	6 ABP79960	Abp79960 N. gonorr
38	135.5	15.9	330	5 ABP38966	Abp38966 Staphyloc
39	130.5	15.3	1035	4 ABB61120	Abb61120 Drosophil
40	129	15.1	115	3 AAG34786	Aag34786 Arabidops
41	129	15.1	143	3 AAG34784	Aag34784 Arabidops
42	127.5	15.0	320	2 AAW89744	Aaw89744 Staphyloc
43	127.5	15.0	320	6 ABJ18994	Abj18994 Pathogen
44	127.5	15.0	320	6 ABW71095	Abw71095 Staphyloc
45	123	14.4	1082	4 ABB63355	Abb63355 Drosophil

ALIGNMENTS

RESULT 1
AAW18312
ID AAW18312 standard; protein; 163 AA.
XX
AC AAW18312;
XX
DT 09-DSC-1997 (first entry)
XX
DE NIMA-interacting protein Pin1.
XX
KW Pin1; protein interacting with NIMA; cell proliferation; mitosis;
KW peptidyl-propyl cis/trans isomerase; adenocarcinoma; cancer; leukaemia;
KW psoriasis; pemphigus vulgaris; rheumatoid arthritis;
KW acute respiratory distress syndrome; septic shock; inflammation; therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 4..43
FT /label= WW
FT /note= "Conserved tryptophan domain thought to be mediate
protein-protein interactions"
FT Peptide 54..69
FT /note= "putative nuclear localisation signal"
FT Domain 59..163
FT /label= PPI
FT /note= "peptidyl-propyl cis/trans isomerase domain"
XX
WO9717986-A1.
XX
22-MAY-1997.
XX
28-OCT-1996; 96WO-US017334.
XX
13-NOV-1995; 95US-00555912.
XX
(SALK) SALK INST BIOLOGICAL STUDIES.
XX
PI Hunter T, Lu KP;
XX
WPI; 1997-289057/26.
XX
N-PSDB; AAT68888.
XX
PT Protein, Pin1, interacting with NIMA - used for treating cell
proliferative disorders.
XX
PS Claim 4; Page 49-50; 73pp; English.
XX

CC This human polypeptide sequence comprises Pin1, an 18 kDa protein that
 CC has peptidyl-propyl cis/trans isomerase activity, associates with NIMA
 CC protein kinase, inhibits the mitosis-promoting function of NIMA when
 CC overexpressed, and induces mitotic arrest and nuclear fragmentation when
 CC depleted. Its sequence was deduced from a DNA sequence (AA168886)
 CC identified in an HeLa library using a yeast two-hybrid system. A
 CC recombinant expression vector comprising the DNA sequence and host cells
 CC containing the vector are claimed. Methods are also claimed for
 CC identifying proteins that inhibit the mitosis promoting function of NIMA
 CC protein kinase and for controlling the growth of a cell by reducing Pin1
 CC activity or Pin1 expression using an inhibitor, anti-Pin1 antibody,
 CC antisense nucleotide sequence or ribozyme, or by increasing Pin1 activity
 CC in the presence of an activator or increasing Pin1 expression using an
 CC enhancer. This allows treatment of cell proliferation disorders such
 CC as adenocarcinomas, cancers, psoriasis, pemphigus vulgaris, acute
 CC respiratory distress syndrome, rheumatoid arthritis, septic shock and
 CC inflammation

XX Sequence 163 AA;

Query Match 100.0%; Score 852; DB 2; Length 163;
 Best Local Similarity 100.0%; Pred. No. 3.5e-83;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MADEEKLPGWEKMSRSGRVYFNHITNASQWERPSGSGGKNGGQEPARVCSHL 60
 DB 1 MADEEKLPGWEKMSRSGRVYFNHITNASQWERPSGSGGKNGGQEPARVCSHL 60
 QY 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120
 DB 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120
 QY 121 DLGAFSRGQMKPFEDASPALRTGEMSGPVFTDSGHIILRTE 163
 DB 121 DLGAFSRGQMKPFEDASPALRTGEMSGPVFTDSGHIILRTE 163

RESULT 2

AA066532
 ID AAG66532 standard; protein; 163 AA.

XX AAG66532;

XX 22-OCT-2001 (first entry)

XX Human interferon-alpha induced polypeptide, PIN-1.

XX Human; interferon-alpha induced gene; type I interferon treatment;
 KW chronic viral hepatitis; relapsing remitting multiple sclerosis;
 KW neoplastic disease; PIN-1.

XX Homo sapiens.

XX WO200159155-A2.

XX 16-AUG-2001.

XX 09-FEB-2001; 2001WO-GB000578.

XX 11-FEB-2000; 2000GB-00003203.

XX 11-FEB-2000; 2000GB-00003204.

XX 11-FEB-2000; 2000GB-00003205.

XX 11-FEB-2000; 2000GB-00003206.

XX 11-FEB-2000; 2000GB-00003207.

XX 11-FEB-2000; 2000GB-00003208.

XX 11-FEB-2000; 2000GB-00003210.

XX 11-FEB-2000; 2000GB-00003212.

XX 11-FEB-2000; 2000GB-00003213.

XX 11-FEB-2000; 2000GB-00003215.

XX 11-FEB-2000; 2000GB-00003216.

PR 11-FEB-2000; 2000GB-00003222.

PR 17-FEB-2000; 2000GB-00003768.

XX (PHAR-) PHARMA PACIFIC PTY LTD.

XX Meritet J, Dron M, Tovey MG;

XX WPI: 2001-483570/52.

DR N-PSDB; AAH76459.

XX Predicting responsiveness of a patient to treatment with a type I

PT interferon comprising determining the level of induced proteins after

PT treatment with a type I interferon.

XX Claim 1; Page 61-62; 133pp; English.

CC The invention relates to a method for predicting responsiveness of a

CC patient to treatment with a type I interferon. The method comprises

CC determining the level of one or more proteins with a 646, 164, 126, 598,

CC 98, 177, 761, 361, 941, 657, 817, 429, 473, 399, 285 or 303 amino acid

CC sequence fully defined in the specification after treatment with a type I

CC interferon. The method allows a physician to determine whether a patient

CC suffering from chronic viral hepatitis, neoplastic disease or relapsing

CC remitting multiple sclerosis will respond favourably to type I interferon

CC treatment via oromucosal administration. The present sequence is one of

CC the polypeptides listed above that may be used in the method

XX Sequence 163 AA;

Query Match 100.0%; Score 852; DB 4; Length 163;

Best Local Similarity 100.0%; Pred. No. 3.5e-83;

Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MADEEKLPGWEKMSRSGRVYFNHITNASQWERPSGSGGKNGGQEPARVCSHL 60

DB 1 MADEEKLPGWEKMSRSGRVYFNHITNASQWERPSGSGGKNGGQEPARVCSHL 60

QY 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120

DB 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120

QY 121 DLGAFSRGQMKPFEDASPALRTGEMSGPVFTDSGHIILRTE 163

DB 121 DLGAFSRGQMKPFEDASPALRTGEMSGPVFTDSGHIILRTE 163

RESULT 3

ADCT2966

ID ADCT2966 standard; protein; 163 AA.

XX ADCT2966;

XX 01-JAN-2004 (first entry)

XX Human petidyl polyl isomerase protein.

XX cytostatic; gynaecological; antipsoriatic; antimicrobial; antibacterial;

XX antiinflammatory; antidiarrhoeic; antipyretic; antirheumatic;

XX immunosuppressive; gastrointestinal; antitubercular; tuberculostatic;

XX fungicide; Pini; petidyl polyl isomerase; PPIase; cancer;

XX fibrotic disorder; non-neoplastic growth; benign hypertrophy;

XX endometriosis; psoriasis; infectious disease; fungal;

XX bacterial infection; pneumonia; diarrhoea; dysentery; anthrax;

XX rheumatic fever; toxic shock syndrome; mastoiditis; meningitis;

XX gonorrhoea; typhoid fever; gastroenteritis; brucellosis; cholera;

XX bubonic plague; tetanus; tuberculosis; Lyme disease; human; enzyme.

XX Homo sapiens.

OS WO2003074002-A2.

XX 12-SEP-2003.

PF 03-MAR-2003; 2003WO-US006518.
 XX
 PR 01-MAR-2002; 2002US-0360799P.
 XX
 PA (PINT-) PINTEX PHARM INC.
 XX
 PI Suto RK, Sowadski J;
 XX
 XX WPI; 2003-779008/73.
 XX
 XX Treating peptidyl prolyl isomerase Pin1 associated disorder e.g. cancer,
 PT infectious diseases and psoriasis comprises administering Pin1 activity
 PT inhibitors.
 PT
 PS Disclosure; SEQ ID NO 1; 4lpp; English.
 XX
 XX The invention relates to a novel method of treating a Pin1 (petidyl polyI
 CC isomerase; ppiase)-associated disorder comprising administration of Pin1
 CC activity inhibitors. The method of the invention may be useful for
 CC treating disorders involving mitosis and cell proliferation, particularly
 CC cancers, fibrotic disorders and non-neoplastic growths including benign
 CC hypertrophy, endometriosis and psoriasis, as well as infectious diseases
 CC including fungal and bacterial infections such as pneumonia, diarrhoea,
 CC dysentery, anthrax, rheumatic fever, toxic shock syndrome, mastoiditis,
 CC meningitis, gonorrhoea, typhoid fever, gastroenteritis, brucellosis,
 CC cholera, bubonic plague, tetanus, tuberculosis and Lyme disease. The
 CC current sequence is that of the human Pin1 protein of the invention.
 XX
 XX Sequence 163 AA;
 SQ
 Query Match 100.0%; Score 852; DB 7; Length 163;
 Best Local Similarity 100.0%; Pred. No. 3.5e-83;
 Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MADEKLPQGWKEMRSRSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60
 Db 1 MADEKLPQGWKEMRSRSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60
 QY 61 LVKHSQRRPSSWRQEKITRTKEALELINGYIQIKSGEDEFSLASQFSDCSAKARG 120
 Db 61 LVKHSQRRPSSWRQEKITRTKEALELINGYIQIKSGEDEFSLASQFSDCSAKARG 120
 QY 121 DLGAFSRGQMKPPEDASFALRTGEMSGPVFTDSGIHILRTE 163
 Db 121 DLGAFSRGQMKPPEDASFALRTGEMSGPVFTDSGIHILRTE 163
 RESULT 4
 AAY48377
 XX AAY48377 standard; protein; 163 AA.
 AC AAY48377;
 XX
 XX 08-DEC-1999 (first entry)
 DT Human prostate cancer-associated protein 74.
 DE Expressed sequence tag; EST; prostate; tumor; treatment; gene therapy;
 KW cancer; tissue specificity; human.
 XX
 XX Homo sapiens.
 OS
 XX DE19811194-A1.
 PN
 XX 16-SEP-1999.
 PD
 XX 10-MAR-1998; 98DE-01011194.
 PF
 XX 10-MAR-1998; 98DE-01011194.
 PR
 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
 PA
 XX Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;

XX WPI; 1999-519629/44.
 DR N-PSDB; AAZ33510.
 XX
 PT New nucleic acid expressed at high level in normal prostatic tissue and
 PT encoded polypeptides, used to treat cancer and screen for therapeutic
 PT agents.
 XX
 PS Claim 25; 152; 194pp; German.
 XX
 XX This invention describes novel nucleic acid sequences (A) that are
 CC expressed at high level in normal prostatic tissue. Polypeptides (I)
 CC encoded by (A) are used: (a) for identifying agents for treatment of
 CC prostatic cancer and (b) for therapy of prostate cancer, optionally where
 CC expressed by gene therapy methods. (A) is also used to isolate full-
 CC length genes (for gene therapy) and for recombinant production of (I),
 CC which can be used to raise specific antibodies. (A) are identified by
 CC assembly of ESTs (expressed sequence tags) before these are analyzed for
 CC expression pattern (tissue specificity). This approach eliminates many of
 CC the false results, as regards tissue specificity, associated with known
 CC methods that use single (usually short) ESTs. AAY48304-Y48456 represent
 CC peptides encoded by the expressed sequence tags described in the method
 CC of the invention
 XX
 SQ Sequence 163 AA;
 Query Match 86.2%; Score 734; DB 2; Length 163;
 Best Local Similarity 97.2%; Pred. No. 1.7e-70;
 Matches 140; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MADEKLPQGWKEMRSRSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60
 Db 1 MADEKLPQGWKEMRSRSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60
 QY 61 LVKHSQRRPSSWRQEKITRTKEALELINGYIQIKSGEDEFSLASQFSDCSAKARG 120
 Db 61 LVKHSQRRPSSWRQEKITRTKEALELINGYIQIKSGEDEFSLASQFSDCSAKARG 120
 QY 121 DLGAFSRGQMKPPEDASFALRTG 144
 Db 121 DLGAFSRGQMKPPEDASFALRTG 144
 RESULT 5
 ABG12572
 ID ABG12572 standard; protein; 191 AA.
 XX
 AC ABG12572;
 XX
 XX 18-FEB-2002 (first entry)
 DT Novel human diagnostic protein #12563.
 DE
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 PN
 XX 11-OCT-2001.
 PD
 XX 30-MAR-2001; 2001WO-US008631.
 PF
 XX 31-MAR-2000; 2000US-00540217.
 PR
 XX 23-AUG-2000; 2000US-00649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX WPI; 2001-639362/73.
 DR N-PSDB; AAS76759.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 42931; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 191 AA;

Query Match 82.7%; Score 705; DB 4; Length 191;
Best Local Similarity 82.6%; Pred. No. 2.8e-67;
Matches 142; Conservative 2; Mismatches 8; Indels 20; Gaps 3;

QY 9 PGWEXMSRSSG-----RVYFNFHITNASQWRPSSGSGGKNGGQ 50
DB 1 PGWEXMSRSSGVTQEAIPDTAAIPDQKRVYFNFHITNASQWRPSSGSGGKNGGQ 60
QY 51 EPARVRCSHLLVKHSQSRPSSWRQEKITRTKEALELINGYIQTKSGEEDFESLASOF 110
DB 61 EPARVRCSHLLVKHSQSRPSSWRQEKITRTKEALELINGYIQTKSGEEDFESLASOF 120
QY 111 SDCSSAKARGDLGAFSGQWQKPF-EDASPALRT-GEMSGPVFTDSGIHIL 160
DB 121 SDCSSAKARGDLGFSRGQWQKPFRTPRFALRTGDERGFCFTDSGIHIL 172

RESULT 6
ABG11947
ID ABG11947 standard; protein; 259 AA.
XX
XX ABG11947;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #11938.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
XX
XX 23-AUG-2000; 2000US-00649167.
XX

PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS76134.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 42306; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 259 AA;

Query Match 75.4%; Score 642; DB 4; Length 259;
Best Local Similarity 76.4%; Pred. No. 2.5e-60;
Matches 133; Conservative 5; Mismatches 14; Indels 22; Gaps 5;

QY 9 PGWEXMSRSS-----GRVYFNFHITNASQWRPSSGSGGKNGGQ 50
DB 1 PGWEXMSRSSVVNTQEAIPDTAAIPDQKRVYFNFHITNASQWRPSSGSGGKNGGQ 60
QY 51 EPARVRCSHLLVKHSQSRPSSWRQEKITRTKEALELINGYIQTKSGEEDFESLASOF 110
DB 61 EPARVRCSHLLVKHSQSRPSSWRQEKITRTKEALELINGYIQTKSGEEDFESLASOF 120
QY 111 SDCSSAKARGDLGAF-SRGQWQKPF-EDASPALRT-GEMSGPVFTDSGIHIL 160
DB 121 SDCCKHPKARGDLGAFQKVRQKPFRTPRFALADGGKXAGCFCTDSGIHIL 174

RESULT 7
AAU32052
ID AAU32052 standard; protein; 195 AA.
XX
XX AAU32052;
XX
XX 18-DEC-2001 (first entry)
XX
XX Novel human secreted protein #2543.
XX
XX Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
XX Homo sapiens.
OS
XX WO200179449-A2.
XX


```

PR 09-JUL-2001; 2001US-0303899P.
PR 31-AUG-2001; 2001US-0316362P.
XX (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX
XX New purified or isolated nucleic acids of essential genes of Aspergillus
XX fumigatus, useful for treating or preventing infections by A. fumigatus,
XX or for treating a non-infectious disease in a subject e.g. cancer.
XX
XX Disclosure; Page; 175pp; English.
XX
XX The invention relates to novel purified or isolated nucleic acids of
XX essential genes of Aspergillus fumigatus. The isolated nucleic acids of
XX the invention are used to treat or prevent infections by a pathogenic
XX organism such as A. fumigatus, to treat a non-infectious disease in a
XX subject (e.g. cancer), to prevent or inhibit contamination of an object
XX by A. fumigatus, or to prevent or inhibit formation on a surface of a
XX biofilm comprising A. fumigatus. The polynucleotides are useful for
XX expressing recombinant protein for characterisation, screening or
XX potential orthologous essential or virulence genes, for selecting and
XX making oligomers for attachment to a nucleic acid array for examination
XX of expression patterns, for raising anti-protein antibodies, as an
XX antigen to raise anti-DNA antibodies or to elicit another immune
XX response, and for identifying polynucleotides encoding the other protein
XX with which binding occurs or to identify inhibitors of the binding
XX interaction. The polypeptides may be used to raise antibodies or to
XX determine levels of the protein in biological fluids, as a marker for
XX host tissues in which pathogenic organism invade or reside, and to
XX isolate correlative receptors or ligands in the case of virulence
XX factors. This sequence represents a protein of one of the essential genes
XX of Aspergillus fumigatus of the invention
XX
XX Sequence 178 AA;
XX
XX Query Match 51.3%; Score 437.5; DB 6; Length 178;
XX Best Local Similarity 51.5%; Pred. No. 1.4e-38;
XX Matches 88; Conservative 26; Mismatches 40; Indels 17; Gaps 3;
XX
QY 7 LPFGWEKRMRSRGVYFNFHITNASQWERPSG-----NSSGGK---NQQG 50
DB 9 LPAGWEVHRHNSKNLPYFNFSTKESRWEPPSGTDTLLKVMYMHSAAPAGPDGTAQG 68
QY 51 EPARVRCSHLLVKHSQSRPSSWQEKITRTKEALELINGYIQIKSGEDFESLASQF 110
DB 69 E-GKIRCSHLLIKHRDSRRPSSWFEAEITRSKEAEIILRGHEQIRSGVSLGDIAVSE 127
QY 111 SDCCSSAKARGDLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGIHILR 161
DB 128 SDCCSARKKGLDGFGRGEMQKEFEDAAAFALQPGQVSGIVETASGVHLIER 178
XX
RESULT 10
ABJ26385
ID ABJ26385 standard; protein; 178 AA.
XX
AC ABJ26385;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #1043.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.

```

```

XX OS Aspergillus fumigatus.
XX PN WO200286090-A2.
XX PD 31-OCT-2002.
XX PF 23-APR-2002; 2002WO-US013142.
XX PR 23-APR-2001; 2001US-0285697P.
XX PR 27-APR-2001; 2001US-0287066P.
XX PR 05-JUN-2001; 2001US-0295890P.
XX PR 09-JUL-2001; 2001US-0303899P.
XX PR 31-AUG-2001; 2001US-0316362P.
XX PA (ELIT-) ELITRA PHARM INC.
XX
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX WPI; 2003-093124/08.
XX
XX New purified or isolated nucleic acids of essential genes of Aspergillus
XX fumigatus, useful for treating or preventing infections by A. fumigatus,
XX or for treating a non-infectious disease in a subject e.g. cancer.
XX
XX Disclosure; Page; 175pp; English.
XX
XX The invention relates to novel purified or isolated nucleic acids of
XX essential genes of Aspergillus fumigatus. The isolated nucleic acids of
XX the invention are used to treat or prevent infections by a pathogenic
XX organism such as A. fumigatus, to treat a non-infectious disease in a
XX subject (e.g. cancer), to prevent or inhibit contamination of an object
XX by A. fumigatus, or to prevent or inhibit formation on a surface of a
XX biofilm comprising A. fumigatus. The polynucleotides are useful for
XX expressing recombinant protein for characterisation, screening or
XX therapeutic use, as markers for host tissues in which the pathogenic
XX organisms invade or reside, for comparing with the DNA sequence of A.
XX fumigatus to identify duplicated genes or paralogues having the same or
XX similar biochemical activity and/or function, for comparing with DNA
XX sequences of other related or distant pathogenic organisms to identify
XX potential orthologous essential or virulence genes, for selecting and
XX making oligomers for attachment to a nucleic acid array for examination
XX of expression patterns, for raising anti-protein antibodies, as an
XX antigen to raise anti-DNA antibodies or to elicit another immune
XX response, and for identifying polynucleotides encoding the other protein
XX with which binding occurs or to identify inhibitors of the binding
XX interaction. The polypeptides may be used to raise antibodies or to
XX elicit immune response, as a reagent in assays designed to quantitatively
XX determine levels of the protein in biological fluids, as a marker for
XX host tissues in which pathogenic organism invade or reside, and to
XX isolate correlative receptors or ligands in the case of virulence
XX factors. This sequence represents a protein of one of the essential genes
XX of Aspergillus fumigatus of the invention
XX
XX Sequence 178 AA;
XX
XX Query Match 51.0%; Score 434.5; DB 6; Length 178;
XX Best Local Similarity 50.6%; Pred. No. 3e-38;
XX Matches 88; Conservative 26; Mismatches 43; Indels 17; Gaps 3;
XX
QY 4 EEKLPPGWEKRMRSRGVYFNFHITNASQWERPSG-----NSSGGK---N 47
DB 6 ETGLPAGWEVHRHNSKNLPYFNFSTKESRWEPPSGTDTLLKVMYMHSAAPAGPDGT 65
QY 48 GQGEPAVRCSHLLVKHSQSRPSSWQEKITRTKEALELINGYIQIKSGEDFESLA 107
DB 66 AQGE-GKIRCSHLLIKHRDSRRPSSWFEAEITRSKEAEIILRGHEQIRSGVSLGDIA 124
QY 108 SQFSDCCSARKARGDLGAFSRGQMKPFEDASFALRTGEMSGPVFTDSGIHILR 161
DB 125 VSESDCCSARKKGLDGFGRGEMQKEFEDAAAFALQPGQVSGIVETASGVHLIER 178

```

```

RESULT 11
AAB87657
ID AAB87657 standard; protein; 87 AA.
XX
AC AAB87657;
XX
15-MAY-2001 (first entry)
XX
Bovine mammary tissue derived protein #48.
XX
Bovine; mammary gland; cancer; tumour; angiogenesis.
XX
Bos taurus.
XX
WO200114553-A1.
XX
01-MAR-2001.
XX
23-AUG-2000; 2000WO-NZ000166.
XX
23-AUG-1999; 99US-0150330P.
XX
(GENE-) GENESIS RES & DEV CORP LTD.
PA (NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.
XX
Havukkala IJ, Gleen M, Grigor MR, Molenaar AJ;
XX
WPI; 2001-226619/23.
XX
New polypeptides and polynucleotides encoding the polypeptides, which are
PT expressed in bovine mammary gland tissue, useful for stimulating mammary
PT gland growth or function, or inducing differentiation of milk producing
PT cells.
XX
Claim 11; Page 80; 97pp; English.
XX
The present invention relates to proteins derived from bovine mammary
CC gland cells. The invention is useful for stimulating bovine mammary gland
CC cell growth and function, inhibiting the growth of various mammary gland
CC cancer cells, inhibiting angiogenesis and vascularization of tumours, or
CC modulating the growth of blood vessels in a mammal
XX
Sequence 87 AA;
XX
Query Match 47.7%; Score 406; DB 4; Length 87;
Best Local Similarity 96.1%; Pred. No. 1.3e-35;
Matches 74; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MADEEKLPPGWKRMSSRGVYFNHITNASQWERPSGNSGKNGQGEPAVRVCSHL 60
Db 1 MADEEKLPPGWKRMSSRGVYFNHITNASQWERPSGNSGKNGQGEPTVRVCSHL 60

QY 61 LVKHSQRRPSSWRQEK 77
Db 61 LVKHSQRRPSSWRQEK 77

RESULT 12
ABB43821
ID ABB43821 standard; peptide; 70 AA.
XX
AC ABB43821;
XX
04-FEB-2002 (first entry)
XX
Peptide #11327 encoded by human foetal liver single exon probe.
XX
Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
Homo sapiens.
XX
WO200157277-A2.
XX

```

```

PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human fetal liver.
XX
Claim 27; SEQ ID NO 36456; 639pp + Sequence Listing; English.
XX
The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
Sequence 70 AA;
XX
Query Match 43.8%; Score 373; DB 4; Length 70;
Best Local Similarity 100.0%; Pred. No. 3.4e-32;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 RVYVFNHITNASQWERPSGNSGKNGQGEPAVRVCSHLVXHSQRRPSSWRQEK 80
Db 1 RVYVFNHITNASQWERPSGNSGKNGQGEPAVRVCSHLVXHSQRRPSSWRQEK 60

QY 81 TKEEALELIN 90
Db 61 TKEEALELIN 70

RESULT 13
AAM37729
ID AAM37729 standard; protein; 70 AA.
XX
AC AAM37729;
XX
17-OCT-2001 (first entry)
XX
Peptide #11766 encoded by probe for measuring placental gene expression.
XX
Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX
Homo sapiens.
XX
WO200157272-A2.
XX
09-AUG-2001.
XX
30-JAN-2001; 2001WO-US000663.
XX
04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX

```

```

PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-48897/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX Claim 27; SEQ ID NO 37998; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP:
CC see AAI31315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders
XX Sequence 70 AA;
SQ
    Query Match      43.8%; Score 373; DB 4; Length 70;
    Best Local Similarity 100.0%; Pred. No. 3.4e-32;
    Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 RVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHLLVYKHQSRRPSSWRQEKITR 80
DB 1 RVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHLLVYKHQSRRPSSWRQEKITR 60
QY 81 TKEEALIELIN 90
DB 61 TKEEALIELIN 70
RESULT 14
AM77544
ID AAM77544 standard; protein; 70 AA.
XX AC AAM77544;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 37850.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000668.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.

```

```

XX Example 4; SEQ ID NO 37850; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention
XX Sequence 70 AA;
SQ
    Query Match      43.8%; Score 373; DB 4; Length 70;
    Best Local Similarity 100.0%; Pred. No. 3.4e-32;
    Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 21 RVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHLLVYKHQSRRPSSWRQEKITR 80
DB 1 RVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHLLVYKHQSRRPSSWRQEKITR 60
QY 81 TKEEALIELIN 90
DB 61 TKEEALIELIN 70
RESULT 15
AM64786
ID AAM64786 standard; protein; 70 AA.
XX AC AAM64786;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 36891.
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000667.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PA Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX Example 4; SEQ ID NO 36891; 650pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention

```

Tue Jun 29 08:37:06 2004

SQ Sequence 70 AA;
Query Match 43.8%; Score 373; DB 4; Length 70;
Best Local Similarity 100.0%; Pred. No. 3.4e-32;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 21 RYVYFNHITNASQWERPSGNSGCKNGQGEPAFVRCSHLLVKHSGSRPSSWROEKITR 80
Db 1 RYVYFNHITNASQWERPSGNSGCKNGQGEPAFVRCSHLLVKHSGSRPSSWROEKITR 60
Qy 81 TKEALBLIN 90
Db 61 TKEALBLIN 70

Search completed: June 28, 2004, 18:36:17
Job time : 132 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 18:34:06 ; Search time 30 Seconds
(without alignments)
280.501 Million cell updates/sec

Title: AAC50492

Perfect score: 852

Sequence: 1 MADEEKLPPGWEKMRSSG.....GEMSGPVFTDSGHIILRTS 163

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pap.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pap.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pap.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pap.*
5: /cgn2_6/ptodata/2/iaa/PTCUS COMB.pap.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	852	100.0	163	2	US-09-066-074-2
2	852	100.0	163	2	US-08-555-912A-2
3	852	100.0	163	3	US-09-208-804-4
4	852	100.0	163	3	US-08-801-743-4
5	852	100.0	163	4	US-09-275-900-2
6	369.5	43.4	177	4	US-09-507-242-2
7	215	25.2	39	2	US-03-066-074-8
8	215	25.2	39	2	US-08-555-912A-8
9	215	25.2	39	4	US-09-252-404A-33
10	215	25.2	39	4	US-09-275-900-8
11	179.5	21.1	92	4	US-09-408-020-68
12	172	20.2	31	4	US-09-252-404A-1
13	152	17.8	131	3	US-09-208-804-1
14	152	17.8	131	3	US-08-801-743-1
15	152	17.8	131	4	US-09-621-976-7038
16	152	17.8	131	4	US-09-976-594-617
17	138.5	16.3	698	4	US-09-252-991A-30206
18	138	16.2	93	3	US-09-208-804-3
19	138	16.2	93	3	US-08-801-743-3
20	135.5	15.9	330	4	US-09-134-001C-3811
21	135	15.8	110	4	US-09-489-039A-8030
22	130.5	15.3	111	4	US-09-252-991A-20022
23	129	15.1	105	4	US-09-489-039A-8015
24	127.5	15.0	320	4	US-08-956-171E-5192
25	118	13.8	443	4	US-09-328-352-7725
26	117	13.7	369	4	US-09-540-236-3738
27	116	13.6	735	3	US-08-539-205A-2

28 116 13.6 735 4 US-09-392-163A-2 Sequence 2, Appli
29 115.5 13.6 852 2 US-09-070-060-3 Sequence 3, Appli
30 115.5 13.6 852 3 US-09-357-746-3 Sequence 3, Appli
31 115.5 13.6 854 2 US-09-070-060-4 Sequence 4, Appli
32 115.5 13.6 854 3 US-09-357-746-4 Sequence 4, Appli
33 114 13.4 97 4 US-09-252-991A-20492 Sequence 20492, A
34 114 13.4 906 3 US-08-630-916A-48 Sequence 48, Appl
35 113 13.3 106 4 US-09-328-352-5211 Sequence 5211, Ap
36 109.5 12.9 38 3 US-08-630-916A-36 Sequence 36, Appl
37 108.5 12.7 683 3 US-08-630-916A-46 Sequence 46, Appl
38 107 12.6 299 4 US-09-336-115C-24 Sequence 24, Appl
39 105.5 12.4 38 3 US-08-630-916A-32 Sequence 32, Appl
40 105.5 12.4 446 4 US-09-489-039A-12283 Sequence 12283, A
41 105 12.3 834 3 US-08-539-205A-6 Sequence 6, Appli
42 105 12.3 834 4 US-09-392-163A-6 Sequence 6, Appli
43 103 12.1 927 3 US-08-895-601-6 Sequence 6, Appli
44 102 12.0 274 4 US-09-252-991A-20389 Sequence 20389, A
45 101 11.9 766 3 US-08-539-205A-4 Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-066-074-2
; Sequence 2, Application US/09066074
; Patent No. 5952467
; GENERAL INFORMATION:
; APPLICANT: Hunter et al., Tony
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,074
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/555,912
; FILING DATE: 13-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07251/011001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 163 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-066-074-2

Query Match 100.0%; Score 852; DB 2; Length 163;

Best Local Similarity 100.0%; Pred. No. 8e-85;

Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MADEEKLPPGWEKMRSSGGRVYFNNHTNASQWERPSGNSSGGKNGQGEFARVRCSHL 60

Db 1 MADEEKLPPGWEKMRSSGGRVYFNNHTNASQWERPSGNSSGGKNGQGEFARVRCSHL 60

QY 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120
Db 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120
QY 121 DLGAFSRGQMOKPFEDASFAIRTGEMSGPVFTDSDGIHILRTE 163
Db 121 DLGAFSRGQMOKPFEDASFAIRTGEMSGPVFTDSDGIHILRTE 163

RESULT 2

US-08-555-912A-2
; Sequence 2, Application US/08555912A
; Patent No. 5972697
; GENERAL INFORMATION:
; APPLICANT: Hunter et al., Tony
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/555,912A
; FILING DATE: 13-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07251/011001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 163 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-555-912A-2

Query Match 100.0%; Score 852; DB 2; Length 163;
Best Local Similarity 100.0%; Pred. No. 8e-85;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60
Db 1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60
QY 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120
Db 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120
QY 121 DLGAFSRGQMOKPFEDASFAIRTGEMSGPVFTDSDGIHILRTE 163
Db 121 DLGAFSRGQMOKPFEDASFAIRTGEMSGPVFTDSDGIHILRTE 163

RESULT 3

US-09-208-804-4
; Sequence 4, Application US/09208804
; Patent No. 6030826
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: NOVEL HUMAN PARVULIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fast-SEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/208,804
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: 08/801,743
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0217 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:

; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 163 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1332710
; US-09-208-804-4

Query Match 100.0%; Score 852; DB 3; Length 163;
Best Local Similarity 100.0%; Pred. No. 8e-85;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60
Db 1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60
QY 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120
Db 61 LVKHSQSRPSSWRQEKITRTKEALELINGYIQIKSGEEDFESLASQFSDCSSAKARG 120
QY 121 DLGAFSRGQMOKPFEDASFAIRTGEMSGPVFTDSDGIHILRTE 163
Db 121 DLGAFSRGQMOKPFEDASFAIRTGEMSGPVFTDSDGIHILRTE 163

RESULT 4

US-08-801-743-4
; Sequence 4, Application US/08801743
; Patent No. 6037164
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: NOVEL HUMAN PARVULIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible

```

; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/801,743
;   FILING DATE: Herewith
;   CLASSIFICATION: 530
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Billings, Lucy J.
;   REGISTRATION NUMBER: 36,749
;   REFERENCE/DOCKET NUMBER: PF-0217 US
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415-855-0555
;   TELEFAX: 415-845-4166
;   TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 163 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; IMMEDIATE SOURCE:
;   LIBRARY: GenBank
;   CLONE: 1332710
; US-08-801-743-4

Query Match      100.0%; Score 852; DB 3; Length 163;
Best Local Similarity 100.0%; Pred. No. 8e-85;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPARVRCSHL 60
Db      1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPARVRCSHL 60
QY      61 LVKHSQRRPSSRWQEKITRTKEALELINGYIQIKSGEEDFESLASQFDCSSAKARG 120
Db      61 LVKHSQRRPSSRWQEKITRTKEALELINGYIQIKSGEEDFESLASQFDCSSAKARG 120
QY      121 DLGAFSRGQMKPFEDASFALRTGEMSGPVFTDGGIHIILRT 163
Db      121 DLGAFSRGQMKPFEDASFALRTGEMSGPVFTDGGIHIILRT 163

RESULT 5
US-09-275-900-2
; Sequence 2, Application US/09275900
; Patent No. 6596848
; GENERAL INFORMATION:
; APPLICANT: Hunter et al., Tony
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
;   ADDRESS: Fish & Richardson P.C.
;   STREET: 4225 Executive Square, Suite 1400
;   CITY: La Jolla
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 92037
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/275,900
;   FILING DATE: 24-Mar-1999
;   CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/555,912
;   FILING DATE: 13-NOV-1995
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07251/011001
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 619/678-5070
;   TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 163 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-275-900-2

Query Match      100.0%; Score 852; DB 4; Length 163;
Best Local Similarity 100.0%; Pred. No. 8e-85;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPARVRCSHL 60
Db      1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPARVRCSHL 60
QY      61 LVKHSQRRPSSRWQEKITRTKEALELINGYIQIKSGEEDFESLASQFDCSSAKARG 120
Db      61 LVKHSQRRPSSRWQEKITRTKEALELINGYIQIKSGEEDFESLASQFDCSSAKARG 120
QY      121 DLGAFSRGQMKPFEDASFALRTGEMSGPVFTDGGIHIILRT 163
Db      121 DLGAFSRGQMKPFEDASFALRTGEMSGPVFTDGGIHIILRT 163

RESULT 6
US-09-507-242-2
; Sequence 2, Application US/09507242
; Patent No. 6537753
; GENERAL INFORMATION:
; APPLICANT: HANES, Steven D.
; APPLICANT: DEVASAHAYAM, Gina
; APPLICANT: CHATURVEDI, Vishnu
; TITLE OF INVENTION: CAESSI: A CANDIDA ALBICANS GENE, METHODS FOR MAKING AND
;   USING, AND TARGETING IT OR ITS EXPRESSION PRODUCTS FOR
;   TITLE OF INVENTION: ANTIFUNGAL APPLICATIONS
; FILE REFERENCE: 454311-2200.1
; CURRENT APPLICATION NUMBER: US/09/507,242
; CURRENT FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 60/121,246
; PRIOR FILING DATE: 1999-02-23
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 177
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-507-242-2

Query Match      43.4%; Score 369.5; DB 4; Length 177;
Best Local Similarity 43.2%; Pred. No. 2.4e-32;
Matches 73; Conservative 34; Mismatches 49; Indels 13; Gaps 2;

QY      7 LPFGWEKMRSSGRVYFNHITNASQWERPSGN-----SSSGKNGQGEPAR 54
Db      8 LPNWTIRVSRSHNKVEYFLNQSTNESSWDPYGTDKVNLNAYIAKFNKNGYKPLVNEGQ 67
QY      55 VRCSHLLVKHSQRRPSSRWQ-EKIRTKEEALELINGYIQIKSGEEDFESLASQFSDC 113
Db      68 VRVSHLLIKNNQSRKPKSKWSDGISGTRDESQILKKHLERILSGEVLKSELANTESDC 127
QY      114 SSARKAGDLGAFSRGQMKPFEDASFALRTGEMSGPVFTDGGIHIILRT 162
Db      128 SSHDRGDLGFFSKGQMQPPFEAAFNHVGVSNIETNSGVHILQRT 176

```

```
RESULT 7
US-09-066-074-8
; Sequence 8, Application US/09066074
; Patent No. 5952467
; GENERAL INFORMATION:
; APPLICANT: Hunter et al., Tony
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,074
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/555,912
; FILING DATE: 13-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07251/011001
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: PIN1/HUMAN
; US-09-066-074-8
;
; Query Match 25.2%; Score 215; DB 2; Length 39;
; Best Local Similarity 100.0%; Pred. No. 1.9e-16;
; Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EKLPGWEKMRSSGRVYFNHITNASQWERPSGNSSS 43
Db 1 EKLPGWEKMRSSGRVYFNHITNASQWERPSGNSSS 39

RESULT 9
US-09-252-404A-33
; Sequence 33, Application US/09252404A
; Patent No. 6495376
; GENERAL INFORMATION:
; APPLICANT: Kun Ping Lu
; APPLICANT: Xiao Zhen Zhou
; TITLE OF INVENTION: Methods and Compositions for Regulating
; TITLE OF INVENTION: Protein-Protein Interactions
; FILE REFERENCE: 1440.1034-000
; CURRENT APPLICATION NUMBER: US/09/252,404A
; CURRENT FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; TYPE: PRT
; ORGANISM: Homo sapien
; US-09-252-404A-33
;
; Query Match 25.2%; Score 215; DB 4; Length 39;
; Best Local Similarity 100.0%; Pred. No. 1.9e-16;
; Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EKLPGWEKMRSSGRVYFNHITNASQWERPSGNSSS 43
Db 1 EKLPGWEKMRSSGRVYFNHITNASQWERPSGNSSS 39

RESULT 10
US-09-275-900-8
; Sequence 8, Application US/09275900
; Patent No. 6596848
; GENERAL INFORMATION:
; APPLICANT: Hunter et al., Tony
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
;
; Query Match 25.2%; Score 215; DB 4; Length 39;
; Best Local Similarity 100.0%; Pred. No. 1.9e-16;
; Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EKLPGWEKMRSSGRVYFNHITNASQWERPSGNSSS 43
Db 1 EKLPGWEKMRSSGRVYFNHITNASQWERPSGNSSS 39

RESULT 8
US-08-555-912A-8
; Sequence 8, Application US/08555912A
; Patent No. 5972697
; GENERAL INFORMATION:
; APPLICANT: Hunter et al., Tony
; TITLE OF INVENTION: NIMA INTERACTING PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/066,074
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/555,912
; FILING DATE: 13-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07251/011001
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: PIN1/HUMAN
; US-09-066-074-8
;
; Query Match 25.2%; Score 215; DB 2; Length 39;
; Best Local Similarity 100.0%; Pred. No. 1.9e-16;
; Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EKLPGWEKMRSSGRVYFNHITNASQWERPSGNSSS 43
Db 1 EKLPGWEKMRSSGRVYFNHITNASQWERPSGNSSS 39
```


STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/275,900
FILING DATE: 24-Mar-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/555,912
FILING DATE: 13-NOV-1995
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07251/011001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
IMMEDIATE SOURCE:
CLONE: PIN1/HUMAN
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-275-900-8

Query Match 25.2%; Score 215; DB 4; Length 39;
Best Local Similarity 100.0%; Pred. No. 1.9e-16;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 EKLPFGWEKMSRSSGRVYFNHITNASQWERPSGNSS 43
DB 1 EKLPFGWEKMSRSSGRVYFNHITNASQWERPSGNSS 39

RESULT 11
US-09-408-020-68
Sequence 68, Application US/09408020
Patent No. 6632937
GENERAL INFORMATION:
APPLICANT: Swanson, Ronald V.
APPLICANT: Feldman, Robert A.
APPLICANT: Schleper, Christa
TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
FILE REFERENCE: DCRP.002A
CURRENT APPLICATION NUMBER: US/09/408,020
CURRENT FILING DATE: 1999-09-29
PRIOR APPLICATION NUMBER: 60/102,294
PRIOR FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 123
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 68
LENGTH: 92
TYPE: PRT
ORGANISM: Cenarchaeum symbiosum
US-09-408-020-68

Query Match 21.1%; Score 179.5; DB 4; Length 92;
Best Local Similarity 40.4%; Pred. No. 4.4e-12;
Matches 44; Conservative 19; Mismatches 23; Indels 23; Gaps 4;

QY 54 RVRCSHLVXHSQRRSSWFQEKITRKBALELINGYIOKIKSGEDFESLASQFS-D 112
DB 4 KIKCSHLVX-----KQGEALAV-----QERLKAGEK-FGKLAKELSID 41

QY 113 CSSAKARGDLGAFSGCMQKPPEDASFALTGEMSGPVFTDSGIHLIR 161
DB 42 GGSARDGSLGYFGRGMVKPPEDAARLQVGEVSEPVKSEFGYHVIKR 90

RESULT 12
US-09-252-404A-1
Sequence 1, Application US/09252404A
Patent No. 6495376
GENERAL INFORMATION:
APPLICANT: Kun Ping Lu
APPLICANT: Xiao Zhen Zhou
TITLE OF INVENTION: Methods and Compositions for Regulating Protein-Protein Interactions
FILE REFERENCE: 1440.1034-000
CURRENT APPLICATION NUMBER: US/09/252,404A
CURRENT FILING DATE: 1999-02-18
NUMBER OF SEQ ID NOS: 42
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 31
TYPE: PRT
ORGANISM: Homo sapien
US-09-252-404A-1

Query Match 20.2%; Score 172; DB 4; Length 31;
Best Local Similarity 96.8%; Pred. No. 6.5e-12;
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 7 LPFGWEKMSRSSGRVYFNHITNASQWERP 37
DB 1 LPFGWEKMSRSSGRVYFNHITNASQWERP 31

RESULT 13
US-09-208-804-1
Sequence 1, Application US/09208804
Patent No. 6030826
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
TITLE OF INVENTION: NOVEL HUMAN PARVULIN-LIKE PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/208,804
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/801,743
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0217 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 131 amino acids
TYPE: amino acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
US-09-208-804-1
Query Match 17.8%; Score 152; DB 3; Length 131;
Best Local Similarity 34.0%; Pred. No. 6.8e-09;
Matches 51; Conservative 14; Mismatches 33; Indels 52; Gaps 9;

QY 37 PSGNSSSG-----GKNGGEPARVRCSHLLV-KHSQSRFPSSMRQE 76
Db 3 PKGSGSGKAGKGAAGSDSADKKAQPKGGNAVVR--HILCEKHG-----49
QY 77 KITRTKEALELINGYIQIKSGEEDFESLASQFDCSSAKARGDLGAFSRGOMKPPED 136
Db 50 KIM-----EAME-----KLKSGMR-FNEVAAQYSE-DKARQGGDLGWMTRGSMVGPFOE 96
QY 137 ASFALRTGEMSGPVFTDS-----GIHIL 160
Db 97 AAFALPVSGMDKPVFTDPPVTKFGYHIIM 126

RESULT 14
US-08-801-743-1
; Sequence 1, Application US/08801743
; Patent No. 6037164
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: NOVEL HUMAN PARVULIN-LIKE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,743
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0217 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 131 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Consensus
; CLONE: Consensus
US-08-801-743-1
Query Match 17.8%; Score 152; DB 3; Length 131;
Best Local Similarity 34.0%; Pred. No. 6.8e-09;
Matches 51; Conservative 14; Mismatches 33; Indels 52; Gaps 9;
```

```
QY 37 PSGNSSSG-----GKNGGEPARVRCSHLLV-KHSQSRFPSSMRQE 76
Db 3 PKGSGSGKAGKGAAGSDSADKKAQPKGGNAVVR--HILCEKHG-----49
QY 77 KITRTKEALELINGYIQIKSGEEDFESLASQFDCSSAKARGDLGAFSRGOMKPPED 136
Db 50 KIM-----EAME-----KLKSGMR-FNEVAAQYSE-DKARQGGDLGWMTRGSMVGPFOE 96
QY 137 ASFALRTGEMSGPVFTDS-----GIHIL 160
Db 97 AAFALPVSGMDKPVFTDPPVTKFGYHIIM 126

RESULT 15
US-09-621-976-7038
; Sequence 7038, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Joerbert, S.
; TITLE OF INVENTION: ESTS and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.Pm
; SEQ ID NO 7038
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-621-976-7038
Query Match 17.8%; Score 152; DB 4; Length 131;
Best Local Similarity 34.0%; Pred. No. 6.8e-09;
Matches 51; Conservative 14; Mismatches 33; Indels 52; Gaps 9;

QY 37 PSGNSSSG-----GKNGGEPARVRCSHLLV-KHSQSRFPSSMRQE 76
Db 3 PKGSGSGKAGKGAAGSDSADKKAQPKGGNAVVR--HILCEKHG-----49
QY 77 KITRTKEALELINGYIQIKSGEEDFESLASQFDCSSAKARGDLGAFSRGOMKPPED 136
Db 50 KIM-----EAME-----KLKSGMR-FNEVAAQYSE-DKARQGGDLGWMTRGSMVGPFOE 96
QY 137 ASFALRTGEMSGPVFTDS-----GIHIL 160
Db 97 AAFALPVSGMDKPVFTDPPVTKFGYHIIM 126

Search completed: June 28, 2004, 18:40:37
Job time : 31 secs
```

OM nucleic - nucleic search, using sw model

Run on: June 28, 2004, 14:35:24 ; Search time 2080.37 Seconds
(without alignments)
9875.458 Million cell updates/sec

Title: AAC50492

Perfect score: 474

Sequence: 1 ATGGCTGCTACGAGCAAC.....TTCTGGAGAGCGTGTAG 474

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.ov.*

5: gb.pat.*

6: gb.ph.*

7: gb.pl.*

8: gb.pr.*

9: gb.ro.*

10: gb.sts.*

11: gb.sy.*

12: gb.un.*

13: gb.vi.*

14: gb.ba.*

15: em.fun.*

16: em.hum.*

17: em.in.*

18: em.mu.*

19: em.om.*

20: em.or.*

21: em.ov.*

22: em.pat.*

23: em.ph.*

24: em.pl.*

25: em.ro.*

26: em.sts.*

27: em.un.*

28: em.vi.*

29: em.htg.hum.*

30: em.htg.inv.*

31: em.htg.other.*

32: em.htg.mus.*

33: em.htg.pln.*

34: em.htg.rod.*

35: em.htg.mam.*

36: em.htg.vrt.*

37: em.sy.*

38: em.htgo.hum.*

39: em.htgo.mus.*

40: em.htgo.other.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	301.2	63.5	621	8	AY086702	Arabidops
2	233	43.2	411	8	AY097375	Arabidops
3	233	49.2	594	8	AY070749	Arabidops
4	226.2	47.7	617	8	MSU09465	Medicago sa
5	224.6	47.4	660	8	MSU09464	Medicago sa
6	224.6	47.4	679	8	MSU09461	Medicago sa
7	224.6	47.4	693	8	MSU09460	Medicago sa
8	223.4	47.1	102183	8	AC027134	Arabidops
9	223.4	47.1	103157	8	AC011810	Arabidops
10	220.4	46.5	96379	8	ATT5K6	Arabidops
11	220.4	46.5	98473	8	ATT3OR20	Arabidops
12	218.2	46.0	125590	3	CEY49B10	Caenorhabdi
13	214.8	45.3	4534	3	DHHS334	D.hydrei Gen
14	210	44.3	119578	2	AC130200	Medicago
15	208.6	44.0	411	8	AY078027	Arabidops
16	208.6	44.0	706	8	AF385735	Arabidops
17	206	43.5	129547	8	AC006434	Genomic s
18	205.6	43.4	571	8	MSU09462	Medicago sa
19	203	42.8	595	8	MSU09463	Medicago sa
20	203	42.8	723	8	AB015760	Nicotiana
21	202.2	42.7	922	8	BSU09459	Medicago sa
22	200.8	42.4	699	6	AX364448	Sequence
23	200.4	42.3	113431	8	AC119413	Medicago
24	199.8	42.2	132883	2	AC146875	Brassica
25	199	42.0	717	8	AY086668	Arabidops
26	199	42.0	729	6	AX364455	Sequence
27	199	42.0	792	8	BT002391	Arabidops
28	199	42.0	834	8	BT003326	Arabidops
29	198.8	41.9	113431	8	AC119413	Medicago
30	197.4	41.6	826	8	AK120807	Oryza sat
31	197.4	41.6	2339	8	AK100960	Oryza sat
32	195.2	41.2	585	6	AX364443	Sequence
33	194.8	41.1	579	8	MSHS3C3	Medicago sa
34	194.6	41.1	411	8	AY077654	Arabidops
35	194.6	41.1	655	8	AY037250	Arabidops
36	194.6	41.1	70952	8	AB024028	Arabidops
37	194.2	41.0	618	8	MSHS3C1	Medicago sa
38	194.2	41.0	668	8	LEH33	Medicago sa
39	194.2	41.0	1089	8	MSHS311	Arabidops
40	193.4	40.8	411	8	AY081741	Arabidops
41	193.4	40.8	471	8	AF370577	Arabidops
42	193.4	40.8	626	8	AY039904	Arabidops
43	193.4	40.8	669	6	AX364449	Sequence
44	193.4	40.8	682	6	AX364072	Sequence
45	193.4	40.8	682	6	AX366162	Sequence

ALIGNMENTS

RESULT 1

AY086702

LOCUS Arabidopsis thaliana clone 268982 mRNA, complete sequence. PLN 14-APR-2003

DEFINITION Arabidopsis thaliana

ACCESSION AY086702

VERSION AY086702.1

KEYWORDS FLI CDNA..

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 621)

AUTHORS Haas,B.J., Volkovskiy,N., Town,C.D., Troukhan,M., Alexandrov,N., Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.

TITLE Full-length messenger RNA sequences greatly improve genome annotation
 JOURNAL Genome Biol. 3 (6), RESEARCH0029 (2002)
 MEDLINE 22088475
 PUBMED 12093376
 REFERENCE 2 (bases 1 to 621)
 AUTHORS Brover, V., Troukhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and Feldmann, K.

TITLE Full-length cDNA from Arabidopsis thaliana
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 621)
 AUTHORS Brover, V., Troukhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and Feldmann, K.

TITLE Direct Submission
 JOURNAL Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA

COMMENT This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ws or Laer ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genet carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.

FEATURES
 source Location/Qualifiers
 1..621
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="268982"
 92..502
 /codon_start=1
 /product="histone H3 protein, putative"
 /protein_id="AA063756.1"
 /db_xref="GI:21555021"
 /translation="MARTQATARKSHGKAPTKQLATKARKSAPTTGCVKPKHPRP
 GFVALREIRYQKSTELLNRKLPFQLRVREIADQFDLRFQSHAVLAQBAEAYLV
 GLFEDNLCAIHAKRVTIMPKDVLARRAERA"

ORIGIN
 Query Match 63.5%; Score 301.2; DB 8; Length 621;
 Best Local Similarity 81.9%; Pred. No. 5.6e-85;
 Matches 389; Conservative 0; Mismatches 23; Indels 63; Gaps 1;

QY 1 ATGGCTCGTACGAGCAAACTGCAAGAAAATCACCGGAGAAAAGTCCGAGGACTCTG 60
 DB 92 ATGGCTCGTACGAGCAAACTGCAAGAAAATCACCGGAGAAAAGTCCGAGGAGCAG 151

QY 61 CTGCTACCAAGGGCGGAGGAATCTGCGCGACTACTCGAGGAGTCAGAACTCCAC 120
 DB 152 CTGCTACCAAGGGCGGAGGAATCTGCGCGACTACTCGAGGAGTCAGAACTCCAC 211

QY 121 CGTTACCGTCCCGAACCGCTCGCTCTCTGAGATTGTAATACCAAGAGACACAGAG 180
 DB 212 CGTTACCGTCCCGAACCGCTCGCTCTCTGAGATTGTAATACCAAGAGACACAGAG 271

QY 181 TTGTTGATCGGTAAACTTCCTTTTCAACGCTGTTCTGCGTAATCGCTCAAGATTACAAG 240
 DB 272 TTGTTGACCGGTAAACTTCCTTCCAACTGTTCTGCGTAATCGCTCAAGATTCAA- 330

QY 241 ACGGATCTGAGATTCAGAGCCATCGCGTGTAGCTCTTCAAGAGCTGCTGAGCATAT 300
 DB 331 ----- 330
 QY 301 TTGACGGATCTGAGATTCAGAGCCATCGCGTGTAGCTCTTCAAGAGCTGCGGAGCA 360

Db 331 --GACGGATCTGAGATTCCAAAGCCATCGAGTGTAGCTCTTCAAGAGCTCGGAGCA 388
 QY 361 TATTTGGTGGGTTTGTGTTGAAGACACAAATCTGTGCCATTATGCAAGAGGTTTACG 420
 Db 389 TATTTGGTGGGTTTGTGTTGAAGACACAAATCTCTGTGCCATTATGCAAGAGGTTTACC 448
 QY 421 ATAAATGCTTAAAGATGTTCAATGGCAAGAGGATTGTTGGAGAGCGTGTAG 474
 Db 449 ATTATGCTTAAAGATGTTCAATGGCAAGAGGATTGTTGGAGAGCGTGTAG 502

RESULT 2
 LOCUS AY097375 411 bp mRNA linear PLN 05-MAY-2002
 DEFINITION Arabidopsis thaliana AT5g10980/T30N20_250 mRNA, complete cds.
 ACCESSION AY097375
 VERSION AY097375.1 GI:20453304
 KEYWORDS FLI CDNA
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE
 1 (bases 1 to 411)
 Shinn, P., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Banh, J., Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.
 Arabidopsis ORF clones

TITLE Arabidopsis ORF clones
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 411)
 AUTHORS Shinn, P., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Banh, J., Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

TITLE Direct Submission
 JOURNAL Submitted (15-APR-2002) Salk Institute Genomic Analysis Laboratory (SIGNAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGRF (SSP) Consortium members constructed and sequenced the PUNI (ORF) clones using the RAFL cDNAs: Shinn, P., Chen, H., Cheuk, R., Kim, C.J., Meyers, M.C., Banh, J., Bowser, L., Chang, E., Dale, J.M., Goldsmith, A.D., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W., Theologis, A., and Ecker, J.R.

Shinn, P. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.

FEATURES
 source Location/Qualifiers
 1..411
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /chromosome="5"

```

/clone="U01392"
/ecotype="Columbia"
/note="this clone is in pUNI 51"
1. .411
/note="histone H3 protein"
/codon_start=1
/product="AT5g10980/T3ON20_250"
/protein_id="AAMI9891.1"
/db_xref="GI:20453305"
/translation="MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRP
GTVALRIRYKQSTELLIRKLPFQRLVREIAODFKTLRFQSHAVLALQEAABAYLV
GLPFDTNLCALHAKRVTMPKDIQLARRIGERA"

ORIGIN
Query Match 49.2%; Score 233; DB 8; Length 411;
Best Local Similarity 72.9%; Pred. No. 3.8e-63;
Matches 345; Conservative 0; Mismatches 65; Indels 63; Gaps 1;

QY 1 ATGGCTCGTACGAACAACTGCAAGAAATACACACGAGGAGAAAGCTCCGAGGACTCTG 60
DB |||||
QY 61 CTCGCTACCAAGGGCGGAGGAATCTGCGCGACTACTGAGGAGTCAAGAAACCTCAC 120
DB |||||
QY 61 CTTGCCACCAAGGGCGGAGGAATCAGCAACCAACCGGAGGAGTCAAGAGCCTCAC 120
DB |||||
QY 121 CGTTACGCTCCGGAACCGTCTGCTTCTGAGATTCGTAATACCAAGAGCACAGAG 180
DB |||||
QY 121 CGTTACGCTCCGGAACCGTCTGCTTCTGAGATTCGTAATACCAAGAGTACTGAA 180
DB |||||
QY 181 TTGTTGATCCGTAACCTCTTTCAAGCTCTGTTCTGTAATCGCTCAAGATTACAA 240
DB |||||
QY 181 TTGTTGATCCGTAACCTCTTCCGAGCTCTGTTCTGGAATAGCTCAGACTTCA 240
DB |||||
QY 241 ACGGATCTGAGATTCAGAGCCATGCGGTGTTAGCTCTTCAAGAGCTGCTCAAGCAT 300
DB |||||
QY 241 ACGGATCTGAGGTTTCAGAGTCAAGGCTGTTAGCTCTTCAAGAGCGGCT----- 291
DB |||||
QY 301 TTGAGGATCTGAGATTCAGAGCCATGCGGTGTTAGCTCTTCAAGAGCTGCGGAGCA 360
DB |||||
QY 361 TATTGTTGGGTTGTTTGAAGACAAATCTGTGCTGCAATCATGCAAGAGGTTTACG 420
DB |||||
QY 298 TACCTAGTTGACTTTTCGAGATTAATCTCTGCTATTCAGCCCAAGAGGTTACG 357
DB |||||
QY 421 ATAATGCTTAAGATGTTCAATGTCAGAGGATTCGTGAGAGCGGTGCTTA 473
DB |||||
QY 358 ATCATGCTTAAGACATCAATGCTAGAGAAATTCGTGAGAACGTGCTTA 410
DB |||||

RESULT 3
AY070749 594 bp mRNA linear PLN 26-DEC-2001
LOCUS Arabidopsis thaliana AT5g10980/T3ON20_250 mRNA, complete cds.
DEFINITION Arabidopsis thaliana (thale cress)
ACCESSION AY070749
VERSION AY070749.1 GI:17979504
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE 1 (bases 1 to 594)
AUTHORS Cheuk, R., Chen, H., Kim, C.J., Meyers, M.C., Banh, J., Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.
TITLE Arabidopsis cDNA clones

Arabisidopsis cDNA clones
```

JOURNAL
REFERENCEUnpublished
2 (bases 1 to 594)

Cheuk, R., Chen, H., Kim, C.J., Meyers, M.C., Banh, J., Bowser, L., Carninci, P., Chang, E., Dale, J.M., Goldsmith, A.D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J.M., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

Direct Submission

Submitted (19-DEC-2001) Salk Institute Genomic Analysis Laboratory (SIGAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA

COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-length cDNA': Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.).

The Salk, Stanford, POBC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Cheuk, R., Chen, H., Kim, C.J., Meyers, M.C., Shinn, P., Banh, J., Bowser, L., Dale, J.M., Goldsmith, A.D., Jiang, P.X., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Liu, S.X., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Pham, P.K., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Davis, R.W., Theologis, A., and Ecker, J.R.

Cheuk, R. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.

FEATURES

Location/Qualifiers

source

1. .594
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAFL06-13-E17(R21392)"
/ecotype="Columbia"
1. .92

5'UTR

93. .503
/note="histone H3 protein"
/codon_start=1
/product="AT5g10980/T3ON20_250"
/protein_id="AALS0088.1"
/db_xref="GI:17979505"
/translation="MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRYRP
GTVALRIRYKQSTELLIRKLPFQRLVREIAODFKTLRFQSHAVLALQEAABAYLV
GLPFDTNLCALHAKRVTMPKDIQLARRIGERA"
504. .594

3'UTR

ORIGIN

Query Match 49.2%; Score 233; DB 8; Length 594;
Best Local Similarity 72.9%; Pred. No. 4.1e-63;
Matches 345; Conservative 0; Mismatches 65; Indels 63; Gaps 1;

QY 1 ATGGCTCGTACGAACAACTGCAAGAAATACACACGAGGAGAAAGCTCCGAGGACTCTG 60
DB |||||
QY 93 ATGGCTCGTACGAACAAACCGCTAGGAAATCCACCGGAGTAAAGCTCCCGAAGCAG 152
DB |||||
QY 61 CTGCTACCAAGGGCGGAGGAATCTGCGCGACTACTGAGGAGTCAAGAAACCTCAC 120
DB |||||
QY 153 CTGCCCACCAAGGGCGGAGGAATCAGACCAACCGGAGGAGTCAAGAGCTCAC 212
DB |||||
QY 121 CGTTACCGTCCCGAACCGTCTGCTTCTGAGATTCGTAATACCAAGAGCACAGAG 180
DB |||||
QY 213 CGTTACCGTCCCGAACCGTCTGCTTCTGAGATTCGTAAGTACCAAGAGTACTGAA 272
DB |||||
QY 181 TTGTTGATCCGTAACCTCTTTCAAGCTCTGTTCTGTAATCGCTCAAGATTACAA 240
DB |||||
QY 273 TTGTTGATCCGCAAGCTTCCCTTCCAGCGTCTCGTTCGGGAAATAGCTCAGGACTTCAAG 332
DB |||||

```

241 ACGGATCTGAGATTCAGAGCCATCGGCTGTAGCTCTTCAAGAAGCTGCTGAAGCATAT 300
333 ACGGATCTGAGGTTTCAGAGTCAAGGCTGCTGTAGCTCTTCAAGAAGCTGCTGCTGCT 383
301 TTGAGGATCTGAGATTCAGAGCCATCGGCTGTAGCTCTTCAAGAAGCTGCTGCTGCT 360
384 -----GAGGCA 389
361 TATTTGGTGGTTTGTTCGAAGACAAATCTGTGTCATTCATCAAGAGGCTTACG 420
390 TACCTAGTTGGACTTTTCAGGATCTATCTCTGCTGCTATTCACGCCAAGAGGCTTACG 449
421 ATAATGCCCTAAGATGTTCAATGTCGAAGAGATTCGTGGAGAGGCTGCTTA 473
450 ATCATGCTCAAGACATTCATTAATTTGGCTAGAGAAATTCGTGGAGAGCTGCTTA 502

RESULT 4
MSU09465 617 bp mRNA linear PLN 23-NOV-1996
LOCUS Medicago sativa Regen S clone p3c131 histone H3.2 mRNA, complete
DEFINITION cds.
ACCESSION U09465
VERSION U09465.1 GI:488576
KEYWORDS Medicago sativa
SOURCE Medicago sativa
ORGANISM Medicago sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE
1 (bases 1 to 617)
Robertson,A.J.
Histone H3 genes in alfalfa
Thesis (1994) Division of Cell Biology and Biophysics, University
of Missouri-Kansas City
2 (bases 1 to 617)
Robertson,A.J., Kepros,T., Dudits,D. and Waterborg,J.H.
Identification of three highly expressed replacement histone H3
genes of alfalfa
DNA Seq. 6 (3), 137-146 (1996)
96293858
MEDLINE 8722568
PUBMED
REFERENCE
3 (bases 1 to 617)
Robertson,A.J.
Direct Submission
Submitted (06-MAY-1994) Anthony J. Robertson, University of
Missouri-Kansas City, Division of Cell Biology and Biophysics, 5100
Rockhill Road, Kansas City, MO 64110, USA
JOURNAL Location/Qualifiers
FEATURES
source
1. .617
/organism="Medicago sativa"
/mol_type="mRNA"
/strain="Regen S"
/db_xref="taxon:3879"
/clone="p3c131"
/clone_lib="LUCA"
/dev_stage="embryo"
<1. .17
/evidence=experimental
18. .428
/codon_start=1
/evidence=experimental
/product="histone H3.2"
/protein_id="AAB36498.1"
/db_xref="GI:488577"
/translaton="MARTKOTARKSTGGKAPRKOLATKAARKSAPTTGGVKPHRYRP
GTVALREIKYQKSTELLIRKLPORLVRIADQFKTLRFQSHAVLALQEAEEAVLV
GLPDTNLCAIHAKRVIMPDKIQLARRINGERA"
21. .425
/mat_peptide
/product="unnamed"
/evidence=experimental
variation
53
/note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replace="t"
242
variation
/note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replace="c"
264
variation
/note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replace="c"
429. .613
3'UTR
/evidence=experimental
449
variation
/note="Sequence variation with genomic H3.2 clone msH3g1,
GenBank Accession Number U09458"
/replace="c"
527. .576
repeat_region
/evidence=experimental
/rpt_type="DIRBCT"
/rpt_unit="527. .551"
597. .603
polya_signal
/evidence=experimental
614
polya_site
/evidence=experimental
ORIGIN
Query Match 47.7%; Score 226.2; DB 8; Length 617;
Best Local Similarity 84.2%; Pred. No. 6.2e-61;
Matches 255; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
QY 1 ATGGCTCTGATCGAAGCAAACTCCAGAAAATCACAAGGAGGAAAAGTCCGAGGACTCTG 60
DB 18 ATGGCTCTGATCGAAGCAAACTCCAGAAAATCACAAGGAGGAAAAGTCCGAGGAGCAG 77
QY 61 CTCGGTCTACCAAGCGCGGAGGAAAATCTGCGCGACTCTCTGAGGAGTCAAGAAAACCTCAC 120
DB 78 CTCGCCACCAAGGCTGCTAGGAAATCTGCTCTACTCTGAGGAGTCAAGAAAACCTCAC 137
QY 121 CGTTACCGTCCCGGAACCGTCTCTCTGAGATTCGTAATACCAAGAGACAGAG 180
DB 138 CGATACCGCCTCGAACTCTGCTCTCTGAGATTCGTAATACCAAGAGAGTACCGAG 197
QY 181 TTGTTGATCGGTAAGTCTCTCTTCAAGCTCTTTTCTGTAATCGCTCAAGATTACAAG 240
DB 198 CTTTGTATCGCAAGCTTCCATTTGAGCTCTTTCGCTGAAATCGCTCAAGATTACAAG 257
QY 241 ACGGATCTGAGATTCAGAGCCATCGGCTGTAGCTCTTCAAGAAGCTGCTGAAGCATAT 300
DB 258 ACGGATTCAGATTCAGAGCCATCGGCTGTAGCTCTTCAAGAAGCTGCTGAAGCTTAC 317
QY 301 TTG 303
DB 318 CTG 320
RESULT 5
MSU09464 660 bp mRNA linear PLN 23-NOV-1996
LOCUS Medicago sativa Regen S clone p3c131 histone H3.2 mRNA, complete
DEFINITION cds.
ACCESSION U09464
VERSION U09464.1 GI:488574
KEYWORDS Medicago sativa
SOURCE Medicago sativa
ORGANISM Medicago sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE
1 (bases 1 to 660)
Robertson,A.J.

```



```

3'UTR      467..693
            /evidence=experimental
variation   487
            /note="Sequence variation with genomic H3.2 clone msh3g1,
            GenBank Accession Number U09458"
            /replace="c"
variation   537
            /note="Sequence variation with genomic H3.2 clone msh3g1,
            GenBank Accession Number U09458"
            /replace="t"
repeat_region 622..645
            /note="Repeat is present as a single sequence element in
            the 3'UTR of the known gene sequence of msh3g1, GenBank
            Accession Number U09458"
            /rpt_type=DIRECT
            /rpt_unit="622..633"
polyA_signal 647..653
            /note="The stem-loop is located among the two
            polyadenylation signals"
polyA_signal 666..672
            /evidence=experimental
polyA_site  693
            /note="20 A residues"
            /evidence=experimental

```

ORIGIN

```

Query Match      47.4%; Score 224.6; DB 8; Length 693;
Best local Similarity 83.8%; Pred. No. 2.1e-60;
Matches 254; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 1 ATGGCTCTGACGAGCAAACTGCAAGAAATACACGAGGAGAAAGTCCGAGGACTCTG 60
   |||||
Db  56 ATGGCTGTGTACAGCAAAACCGCTGCGAAATCCACTGGTGTAGGCTCCAGGAGCAG 115
   |||||

QY 61 CTCGCTACCAAGGGCGGAGAAATCTGCCGACTACTCGAGGAGTCAAGAAACCTCAC 120
   |||||
Db  116 CTCGCCACCAAGGCTGTAGGAATCTGCTCTACTCTGGAGAGTCAAGAAACCTCAC 175
   |||||

QY 121 CGTTACGCTCCGGAACCGTCTGCTGTCAGATTCGTAATACCAAGAGCACAGAG 180
   |||||
Db  176 CGATACCGCTCGAAGCCGCTCTGCTGAGATCCGTAAGTACCAGAGAGTACCGAG 235
   |||||

QY 181 TTGTTGATCCCTAAACTTCCTTTCAACGCTCTGTTCTGTAATCGCTCAAGATTACAAG 240
   |||||
Db  236 CTTTGTATCCGCAAGCTTCCATTTTCAGGCTCTGTGCGTGAATGCTCAAGATTCAAG 295
   |||||

QY 241 ACGGATCTGAGATTCAGAGCCATCGGTGTGTAGCTTTCAGAAAGCTGCTGAAGCATAT 300
   |||||
Db  296 ACGGATTTGAGATTCAGAGCCATCGAGTCTTCTTGCACTTCAGGAAGCAGCTGAGGCTTAC 355
   |||||

QY 301 TTG 303
Db 356 CTG 358

```

```

RESULT 8
AC027134/c
LOCUS      AC027134      102183 bp      DNA      linear      PLN 15-AUG-2000
DEFINITION Arabidopsis thaliana chromosome I BAC F13B4 genomic sequence,
complete sequence.
ACCESSION  AC027134
VERSION     AC027134.4
KEYWORDS   HTG
SOURCE     Arabidopsis thaliana (thale cress)
ORGANISM   Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 102183)
Pederspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Miramanda,M., Brooks,S.,
Buehler,E., Chao,Q., Chin,C., Chiou,J., Choi,E., Dunn,P., Gonzalez,A.,
Howng,B., Kim,C., Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharly,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A., and Davis,R.W.
Direct Submission
Submitted (28-MAR-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
3 (bases 1 to 102183)
Pederspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E.,
Chin,C., Chiou,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C.,
Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharly,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A., and Davis,R.W.
Direct Submission
Submitted (21-JUN-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
4 (bases 1 to 102183)
Pederspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.,
and Davis,R.W.
Direct Submission
Submitted (15-AUG-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
On Jun 21, 2000 this sequence version replaced gi:8134868.
Bases 1-55,850 of F13B4 overlap with bases 47,308-103,157 of BAC
clone T6J4, gb|AC011810 and bases 76,686-102,183 of F13B4 overlap
with bases 1-25,498 of BAC clone F21F23, gb|AC027656.
e-mail for correspondence: arabsequence.stanford.edu
Genes with
similarity to proteins in the databases are named 'putative',
'-like' or 'similar to'. Genes that have EST similarity but no
significant protein similarity are described as 'unknown proteins'.
Genes that are annotated based only on gene prediction software
are described as 'hypothetical proteins'. The gene prediction
programs used to predict genes include: Grail (Informatics Group,
Oak Ridge National Laboratory,
http://combio.ornl.gov/section/index.html), GENSCAN (Chris Burge,
http://genes.mit.edu/GENSCAN.), FEXA (Victor Solovyev,
http://genomic.sanger.ac.uk/gf/gf.shtml), and NetPlantGene (S.M.
Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).
FEATURES             source
1..102183
   /organism="Arabidopsis thaliana"
   /mol_type="genomic DNA"
   /db_xref="taxon:3702"
   /chromosome="I"
   /clones="F13B4"
1..55850
   /note="Overlap with bases 47,308-103,157 of BAC clone
   T6J4, gb|AC011810. See GenBank record for clone T6J4 for
   annotation in this region."
   /complement(55518..56867)
   /gene="F13B4.1"
   /complement(join(55518..56358,56467..56566,56660..56867))
   /gene="F13B4.1"
   /note="Hypothetical protein"
   /codon_start=1
   /protein_id="AAF99819.1"
   /db_xref="GI:9802750"
misc_feature
T6J4, gb|AC011810. See GenBank record for clone T6J4 for
annotation in this region."
gene
CDS

```

```

JOURNAL
REFERENCE
AUTHORS
Hewng,B., Johnson-Hopson,C., Khan,S., Kim,C., Koo,T., Lee,J.M.,
Lenz,C., Liu,A., Liu,S., Mukharly,N., Pham,P., Sakano,H.,
Shinn,P., Toriumi,M., Vaysberg,M., Yu,G., Ecker,J., Theologis,A.,
and Davis,R.W.
Unpublished
2 (bases 1 to 102183)
Pederspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E.,
Chin,C., Chiou,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C.,
Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharly,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A., and Davis,R.W.
Direct Submission
Submitted (28-MAR-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
3 (bases 1 to 102183)
Pederspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E.,
Chin,C., Chiou,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C.,
Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharly,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A., and Davis,R.W.
Direct Submission
Submitted (21-JUN-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
4 (bases 1 to 102183)
Pederspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.,
and Davis,R.W.
Direct Submission
Submitted (15-AUG-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
On Jun 21, 2000 this sequence version replaced gi:8134868.
Bases 1-55,850 of F13B4 overlap with bases 47,308-103,157 of BAC
clone T6J4, gb|AC011810 and bases 76,686-102,183 of F13B4 overlap
with bases 1-25,498 of BAC clone F21F23, gb|AC027656.
e-mail for correspondence: arabsequence.stanford.edu
Genes with
similarity to proteins in the databases are named 'putative',
'-like' or 'similar to'. Genes that have EST similarity but no
significant protein similarity are described as 'unknown proteins'.
Genes that are annotated based only on gene prediction software
are described as 'hypothetical proteins'. The gene prediction
programs used to predict genes include: Grail (Informatics Group,
Oak Ridge National Laboratory,
http://combio.ornl.gov/section/index.html), GENSCAN (Chris Burge,
http://genes.mit.edu/GENSCAN.), FEXA (Victor Solovyev,
http://genomic.sanger.ac.uk/gf/gf.shtml), and NetPlantGene (S.M.
Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).
FEATURES             source
1..102183
   /organism="Arabidopsis thaliana"
   /mol_type="genomic DNA"
   /db_xref="taxon:3702"
   /chromosome="I"
   /clones="F13B4"
1..55850
   /note="Overlap with bases 47,308-103,157 of BAC clone
   T6J4, gb|AC011810. See GenBank record for clone T6J4 for
   annotation in this region."
   /complement(55518..56867)
   /gene="F13B4.1"
   /complement(join(55518..56358,56467..56566,56660..56867))
   /gene="F13B4.1"
   /note="Hypothetical protein"
   /codon_start=1
   /protein_id="AAF99819.1"
   /db_xref="GI:9802750"
misc_feature
T6J4, gb|AC011810. See GenBank record for clone T6J4 for
annotation in this region."
gene
CDS

```

/translation="MYVTKLSEYQKNPSELTSLTEGPNISGLVLIQDEESRTCCFG
 KCFDKMLKPPQNTKTVRFQGVKRVLDVPAFIPVLHQPLSLNLYIQRSGKQT
 GEACVATGEGTSCICFSFIPADAPRDLDFDIYQQFELHQSGSNRYKQFATVAS
 EGIPQFLTKRWGAASLFYEMTGLIDAKGVVDKLRJELFDLKSIVVGVKWWYFPL
 FYKGEQDANDQNKSMYSMTLQORPEEFVFCENTGNKHFWVDVEVEVVKLEG
 DKIAERTKGVNSDGVVWFVSQTKIGLGSVVLERMKWEERFGWSKKDELRSIKRT
 EKFEQGGPHMKSVRCVLTVELTKETDGLSLVTFYFKHVDKLKSKWD"
 complement (58592..61490)
 /gene="F13B4.2"
 /complement (join (58592..59438,59566..59665,59759..59905,
 60390..61084,61171..61490))
 /gene="F13B4.2"
 /note="Hypothetical protein"
 /codon_start=1
 /protein_id="AAF99820.1"
 /db_xref="GI:9802751"
 /translation="MYVTRCLSEYQKNPSELTSLTEGPNISGLVLIQDEESRTCCFG
 GLCLVDSYLSGLPLPQNLKLAARFVAGDDATCDPVFIPVNLHPLSKSNRYAIKRV
 KGHSGAPASGEEDRVPCCFSSVYPEAKPQLDPDMYQOPLHOKKPSRYAYAT
 SVAPDVSFWFLKKNKWTVEYRFQDFELDKDAKGLNTELRTELPLGLNSVTVGKVV
 PFIVKRDVDOIKKSMYSMTLEQREWEVSIEKSKNTDVLVDVDELTEVVKVD
 GGEISRGVANGFVPAVDGKKIGLGSVVRMKWEERFGMTGDNERSMVAKRLE
 KSTDEGPNISGLVLIQDEESRTCKCCFNVSFDWELNGLPFPQNTYIVKRGITGLY
 DVAFIPVLHQPLSLNLYIQRSGKTEGACVSAKSGDRASCFTYIPEAVRPL
 DPFIDNQFELHQSGSTKKFPATSVASDGIPOPLTRKGSVIFSAPEVDYGLIDD
 AKGIVDAKLRLELPOLDNSVVGKVVYVPLFKVEGDADKDMYKSMYSMTLQORPEEV
 FFCENVNKNREVVVDVEVEVVKLEGERIARETQVNSGVVWFNAGSNEKILGL
 SVLTERMKWEERFGWSKKDELRSIKRTERFEGGPIYKSKYRCYVLTVELTKRTD
 GSLVTFYFKHVDKLKSKWD"
 complement (62660..65932)
 /gene="F13B4.3"
 /complement (join (62660..63503,63707..64040,64557..65380,
 65510..65621,65725..65932))
 /gene="F13B4.3"
 /note="Hypothetical protein"
 /codon_start=1
 /protein_id="AAF99821.1"
 /db_xref="GI:9802752"
 /translation="MYVTKLSEYQKNPSELTPLTEGPNISGLVLIQDEESQTCFG
 KCFGDLNGFAPKEEDKVSCLCYYSNATFRLDLPDIYQQFELHQSGSNRYIARR
 KHAGACVSAKEEDKVSCLCYYSNATFRLDLPDIYQQFELHQSGSNRYIARR
 SVASDGIPIFRKRWGAPFVSASQDYGLIDDAKGVVDKLRJELFDLKSIVVGVKWWY
 VPLFYKGEADKQMKSMYSMTLHQREFEVFCENVNKNCEVTVDEVEVVKL
 EGENIGRETGVNSDGVVWFVSQTKIGLGSVVLERMKWEERFGWSKKDELKSSI
 KRTEKEGGQPNKSVRCVLTVELTKETDGLSLVTFYFKHVDKLKSKWYTRH
 LSYORNSDLTSLPEGPNISGLVLIQDEESRTCCFCNCDYDGLKGLPQNAKLT
 TYRTGDNRRSHDQVLTIPVLDPQYNSRYVIKRGKQSGSASAKEDRVFCC
 FCFYVFPKQADFDIYQQFVEHQKRPSSGYITATVAEPNIPPEFLKRYTVE
 YNSQDFGLTDDAKINTLRSKLPNDMNTSVVVGKWWYFIPVPERDAKQIKSPY
 YSMTLQORWEVYSCVNVNKGVSVDVDETQVVKLEGOETISRETSGGFYWF
 SVLGDQRQDKKIGLGSVVRMKWEERFGMLNNGERSIKRSERFEGGSHWSYRC
 LVLIESFELKMDGSLVLTVEFTHVDKLKSKWD"
 complement (66277..69118)
 /gene="F13B4.4"
 /complement (join (66277..67099,67189..67554,67689..67786,
 67968..68724,68802..69118))
 /gene="F13B4.4"
 /note="Unknown protein"
 /codon_start=1
 /protein_id="AAF99822.1"
 /db_xref="GI:9802753"
 /translation="MYVTRLSYRKNPSELKPPPEGPNISGLVLIQDEESLTCCFG
 LCNEIFLGRPLFPQNALTVRRDQGNNTVHRDPVFPVLQPLSLNRYIYRRCY
 YSRGESSSTKEVERVTCFCIRFNHVMKKPQDPVDYHQVVTITSSPSRYC
 FAKVAPNALLPAFLMRQGTIENTSLQDFGLIDDAKGLNLELSKLPGLDLSVV
 KKNVPELFPKEGDIIDQVKLSMYNMTLHQREWEVFFVENVNEBCVEVDVDEL
 QVNIQGVKISITINDDAKGIWVQVKKIKILRSVVRMESESESESEFVGWVDGNR
 VTKRLDFRFGSGSHWSKYCYLSWSVEYFWMQSGVMVQIRGEDFVYLITLDERI
 KFFSKGRTSETMYITRSSEYRRNPAELTPPPEGPNISGLVLIQDEESRTATSCF
 GSWLVDLSLGLPLPQNLKLAFTVIGDDSTRDPVFIPLDKLISNRYVIAIKRH
 KGHSGASGAKEDDVSVCFCSSVYPEAKPQLDPDMYQOPLHOKKPSRYAYAT
 SVAPDVSFWFLKKNKWTVEYRFQDFELDKDAKGLNTELRTELPLGLNSVTVGKVV
 PFIVKEREVDKQIKNSMYISMTLEQREWEVFSDENDKSNRDNVVDVEVEVVKLE

gene

CDS

QQMTERGLEANGFWFVGDKKIGLGSVVRMKWEERFGMTGDNERSMTVVRLE
 KSTGSLWKHCHVYLIESFVKRMDSESLVTFYETHVDKLKTKWD"
 complement (69433..72830)
 /gene="F13B4.5"
 /note="synonym: AAPT1"
 /complement (join (69433..69658,69744..69816,69923..69975,
 70060..70118,70214..70276,70373..70457,70549..70589,
 70670..70733,70849..70918,71026..71094,71224..71318,
 71397..71497,71584..71606,71693..71768,71872..71949,
 72079..72129,72222..72231,72358..72508,72746..72830))
 /gene="F13B4.5"
 /product="AAPT1"
 /complement (join (69637..69658,69744..69816,69923..69975,
 70060..70118,70214..70276,70373..70457,70549..70589,
 70670..70733,70849..70918,71026..71094,71224..71318,
 71397..71497,71584..71606,71693..71768,71872..71949,
 72079..72129,72222..72231,72358..72494))
 /gene="F13B4.5"
 /note="Location of cDNA clone gb|AF091843
 cholinephosphotransferase; ethanolaminophosphotransferase"
 /codon_start=1
 /product="aminoalcoholphosphotransferase"
 /protein_id="AAF99823.1"
 /db_xref="GI:9802754"
 /translation="MGYIGAGVGAALHRVKYSGVDHSYLAKVLOPFWTRFVKVPLM
 WPMNITLMGMFLVTSLSLGVISPOLDSPPRMVHFAHGLLLFLYCTFFDAVDSKQA
 RTNSSPGLBFLDHCALACAFMAFGSTACGROTFNFWLSALIFYGATWEHY
 FNTLILVINGTEGLALIFVSHFFTAIVGAEMWAQGLQSIPIFWSVFFVEIQT
 RAVYLMIAFAVPTVAFNVNVRNRSNWSVLALAMLYPFVLLGGVLIWYLS
 PINLIATYHLVLTGLAFGLVGRMILAHLCDEPKGLKTMCMCLSLYLPALANAL
 TARLNAGVPLVDLGVLCITVSLYHFAVSIHEITELGIYCFRITRKEA"
 complement (73331..74733)
 /gene="F13B4.6"
 /complement (join (73331..73621,73703..73861,73933..74733))
 /gene="F13B4.6"
 /note="Unknown protein"
 /codon_start=1
 /protein_id="AAF99824.1"
 /db_xref="GI:9802755"

gene

CDS

Query Match 47.1%; Score 223.4; DB 8; Length 102183;
 Best Local Similarity 69.2%; Pred. No. 1.1e-59;
 Matches 413; Conservative 0; Mismatches 61; Indels 123; Gaps 3;

QY 1 ATGGCTCGTACGAGCAAACTCCAGAAATCACAGGAGGAAAAGCTCCGAGGACTCTG 60
 DB 19325 ATGGCTCGTACGAGCAAACTCCGAGAAATCACAGGAGGAAAAGCTCCGAGGAG 19266
 QY 61 CTGCTTACCAAGCGCGGAGGAAATCTGCGCGACTACTTGGAGGAGTCAAGAAACCTCAC 120
 DB 19265 CTGCTTACCAAGCGCGGAGGAAATCTGCGCGACTACTTGGAGGAGTCAAGAAACCTCAC 19206
 QY 121 CGTTACCGTCCGGAACCGTCTCTTC-----148
 DB 19205 CGTTTCGCTCGTGAACCGTCTCTTCGTTAACTTTCACGTTTCTTCTTTCGAAATCA 19146
 QY 149 -----GTGAGATTCT 159
 DB 19145 GAATCTTCTCATCGTCTTCAATTTCTGATGACTATTGTTGATGATCATGATGATCTGT 19086
 QY 160 AAATACCAAGAGCAAGAGTTGTTGATCCGTAACCTTCTTTCAAGCTCTTGTCTGT 219
 DB 19085 AAATACCAAGAGCAAGAGTTGTTGATCCGTAACCTTCAAGCTCTTGTCTGT 19026
 QY 220 GAATCGCTCAAGATTACAGACGGA-----TCTGAGATTCCAGAGCATCTC 266
 DB 19025 GAATCGCTCAAGATTACAGAGTAAAGATTGTTCTTTCATCATCGAAACCTTAGAT 18966
 QY 267 GGTGTTAGTCTTCAAGAGCTGCTGAA-----GCA 297
 DB 18965 ACTGATTAGTTATACCTACTTGTATGAAAAAGATTGATAAATTTGGTGGAAATTTGCA 18906
 QY 298 TATTGACGATCTGAGATTCAGAGCCATCGCGTGTAGTCTTCTTCAAGAGCTCGGAG 357


```

/protein_id="AAG09564.1"
/db_xref="GI:9958075"
/translation="MASSVFLSSFSSSSSLSQCSFHEYLAPRCFLGAPVTSSSSLS
LSGKKNSYSPQFHSKAKVSGLSGLBEAIRIRKRELETKSKVRNPPFRGRVSPRLIV
PDHLPYPPYVESGVLPISSFOIPGEGIAKMAACELAAARVLNVAAGTILVKPSVTLV
EIDKAVHDMILIBAGYPSPLGYGFPKSVCTVNECHGCIKPSROLSQSGDILINDVT
VILGYDGDTRTFCGSEVDGPFRLVKVTECLERGLAVCKDGASPKKIKGRISSEA
EKFGYNVVERVHGVSEHSEPLIHYRNDPEGLMVEGQITFIEFLIILITTECVT
WPDNWTILITADGGVAAQFEHTILITRTGSEILTRC"
/gene="T6J4.4"
/join(25951..25968,26138..26662)
/note="T6J4.4"
/gene="Unknown Protein"
/codon_start=1
/protein_id="AAG09557.1"
/db_xref="GI:9958068"
/translation="MTTPDQIQELNVYEFNEGRNSPAVLKLGKKPDLCLGLDLPVFT
NKLYTGDITKRGITAGLCVLIOHVPEKKGRFEASYSFYFGDYGHI SVQGPYLYTED
TFLAITGSGVFEAGYQVKLQQLVPTKLPYTFVLKGVAAADLPVELLTKGKHSKEV
KPAEAQAQTPGATIANFTN"
/complement(28476..29637)
/gene="T6J4.5"
/complement(join(28476..29033,29287..29637))
/gene="T6J4.5"
/note="hypothetical protein, similar to zinc finger
proteins"
/codon_start=1
/protein_id="AAG09553.1"
/db_xref="GI:9958064"
/translation="WYNNNOYFSGDESDVLSLGGPGQYPSHNKPTSTPKPSDHEF
NHPLNPNQVTVLHI GPSSDKETLSGNNQEGLTARQGVWIPSLQILVGPQIFS
CSVKNTNFRNMQHMGHGSQYKGPESLRGTSSSILRLCYCAEGCKONID
HPSKPLKDFRQLQTHKKGAKKPFCKRCKEKTFAVRGDWRKTHEKCKGLKMFVCG
SDFKRSKDKVRAPDGHAAHTSDRVYVGDADEDEDEDEDEDEDEDEDEDEDAHEE
NVRGKNSKIRYDHFRRYQISDDNY"
/gene="T6J4.6"
/join(35087..36701)
/join(35087..35228,35403..35686,35863..36052,36178..36254,
36360..36701)
/gene="T6J4.6"
/note="Unknown Protein"
/codon_start=1
/protein_id="AAG09552.1"
/db_xref="GI:9958063"
/translation="MIKKSNDYQKRCQYIEALEERRKIHVFORELPLCLDL
VTOALEAKRELPETMTENNYQPECEQTGECGPVLEQFLTIKDSSTNSDEEBEF
DDEGNHDPNDSKNTKDWLKSQVLMQNPDLHPKRLQDETMTDRSMRKP
MYNGEGRKREAEKGGGRKQRCWSSQLHRRFLNALQHLGHPHVAFTPKQIREPMKY
DGLTNDVXSHLQKYLHTRPRQTPVNNNGSQTQHFVVVGLVQPSDYSGTKTGG
ATSTSTTTTGTGYGTMAAPPQWPSHSNYPFSIIVDEGSGSHSEGVVVRCSPPAMSS
STRNHVKN"
/complement(39327..41039)
/gene="T6J4.7"
/complement(join(39327..39677,39858..40010,40103..40269,
40350..40590,40669..40836,40995..41039))
/gene="T6J4.7"
/note="Hypothetical Protein"
/codon_start=1
/protein_id="AAG09559.1"
/db_xref="GI:9958070"
/translation="MTMWQYLIPEKPTKEILYEKTLQAAAGPITLGEVKELTSKRRIIE
ESVNTSKVIDSTRMTGRLTSACQDLHLGLEYLPLFLNLVHYTDKIKVSNLKLIR
WSSGLISQTLIQRTCPKFOVDNIMEFGVLFYIAVKLRERAMELVSTDSKAVAVY
REAGVFFHLSHLIELLQCLPPKLPDLTPCLSAISLLCLAEQAVTTEAESGG
KSALESKLSKGLIFQFLSEAYALLSRLTGEYKDLSTFEYVYTTVMGALHELKSKYL
AEILSEDRVGDVGRVLRALAAAKSTPSKDKKIAIIFEKEREDVAKNMAKYEKLND
SNMLKIPIDREIPFPKGEKINLIPYTPRVVRELRFKS"

```

Query Match 47.1%; Score 223.4; DB 8; Length 103157;
 Best local similarity 69.2%; Pred. No. 1.1e-59;
 Matches 413; Conservative 0; Mismatches 61; Indels 123; Gaps 3;

```

QY 1 ATGCTCGTACGAGCAAACCTGCAAGAAATCACACGAGGAGAAAGCTCCGAGACTCTG 60
DB 66332 ATGGCTCGTACGAGCAATCCGAGAAATACACGAGGAGAAAGCTCCGAGAGCAG 66373
QY 61 CTCCTACCAAGCGGCGAGAAATCTGCCGACTACTCGAGGAGTCAAGAAACCTCAC 120
DB 66572 CTCGCTACCAAGCGGCGAGAAATCTGCCGACTACCGGAGGAGTCAAGAAACCTCAC 66513
QY 121 CGTTACGCTCCGAGAAACGTCGCTCTC----- 148
DB 66512 CGTTCCGCTCGTGAACCGTCTCTTCGGTCACTTTCAGCTTTTCTCTTTTTCGAATCA 66453
QY 149 -----GTCGAGATTCTG 159
DB 66452 CAATCTTCTCATCGTCTCTCAATTTCTGATGACTATTGTTGATGAATCAGTGAGATTCTG 66393
QY 160 AAATACCAAGAGACACAGAGTTGATTCGTAACCTTTCCTTTTCAACGCTCTGTTCTG 219
DB 66392 AAATACCAAGAGACACAGAGTTGTTGAACCGTAACTTCCATTCCAAACGCTCTGTTCTG 66333
QY 220 GAATCGCTCAAGATTACAAGACGGA-----TCTGAGATTCCAGAGCCATGC 266
DB 66332 GAATCGCTCAAGATTACAAGTAAAGATTGTTGTTCAATCATCGAAACCTTAGAT 66273
QY 267 GGTGTAGTCTCTCAAGAGCTGCTGAA-----GCA 297
DB 66272 ACTGATTAGTTATACCTAACTTGAATGAAAGATTGATAAATAATTGCTGGAATTTGCA 66213
QY 298 TATTGACGATCTGAGATTCAGAGCCATCGGCTGTTAGTCTTCAAGAGCTCGGAG 357
DB 66212 ATGTAGACGATCTGAGATTCAGAGCCATCGAGTGTAGTCTTCAAGAGCTCGGAG 66153
QY 358 GCATATTTGGTGGTGTCTTTTGAACACAAATCTGTGCAATTCATGCAAGAGGTT 417
DB 66152 GCATATTTGGTGGTGTCTTTTGAACACAAATCTGTGCAATTCATGCAAGAGGTT 66093
QY 418 ACGTAATATGCTTAAAGATTGTTCAATGCAAGAGGATTCGTTGAGAGCGTCTTAG 474
DB 66092 ACCATTATGCTTAAAGATTGTTCAATGCAAGAGGATTCGTTGAGAGCGTCTTAG 66036

RESULT 10
ATT5K6/c
LOCUS
DEFINITION
Arabidopsis thaliana 96379 bp DNA linear PLN 09-AUG-2000
project).
ACCESSION
AL391222.1 GI:9795153
VERSION
AL391222.1
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 96379)
AUTHORS
Bevan, M., Peters, S.A., van Staveren, M., Dirkse, W., Stiekema, W.,
Bancroft, I., Mewes, H.W., Rudd, S., Lemcke, K. and Mayer, K.F.X.
JOURNAL
Unpublished
2 (bases 88399 to 91734)
REFERENCE
Bevan, M., Van Der Schueren, J., Chuang, Y.J., Voet, M., Robben, J.,
Volckaert, G., Bancroft, I., Mewes, H.W., Rudd, S., Lemcke, K. and
Mayer, K.F.X.
AUTHORS
Unpublished
3 (bases 1 to 96379)
JOURNAL
Direct Submission
EU Arabidopsis sequencing, project.
REFERENCE
Submitted (09-AUG-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bbsrc.ac.uk
COMMENT
Information on performance of analysis and a more detailed

```

annotation of this entry and other sequences of chromosomes 3, 4
 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES
 source
 1..96379
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 37579..40123
 /gene="TSK6_10"
 complement (37579..37728,37825..37914,38003..38121,
 38249..38348,38439..38535,38628..38691,38775..38907,
 38990..39169,39247..39386,39988..40123)
 /gene="TSK6_10"
 complement (37579..37728,37825..37914,38003..38121,
 38249..38348,38439..38535,38628..38691,38775..38907,
 38990..39169,39247..39386,39988..40123)
 /gene="TSK6_10"
 notes="similarity to protein kinase 1, Populus nigra,
 EMBL:AB041503
 Contains Prokaryotic membrane lipoprotein lipid attachment
 site AA113-123;Protein kinases signatures and profile
 AA116-138;Protein kinases signatures and profile
 AA231-243"
 /codon_start=1
 /product="ser/thr specific protein kinase-like protein"
 /protein_id="CAC03450.1"
 /db_xref="GI:9795154"
 /db_xref="GOA:Q9FY63"
 /db_xref="SPTREMBL:Q9FY63"
 /translation="MKQIVITLVLLQAYVLHQSCTVMSLTTOESPQPSAFTPAL
 PGSTSAQVNSLSHSFSGFYGLNLEIHDFDKAEKVTNLVSTWLKGVSLIDVNI
 LEGETSGFKNILQGGGFCVYSATLENNISAAVKLDCANEDAAKEFSEVELLSK
 LQENITLSLGYSTNDTARFIVYELMPNVLSHLSHSGQSALTPMPKIALDVTR
 GLEYLHCHPAITHRDLKSNLSNIDSNFNAKISDFGLAVVDGPKNKHLSGTGVY
 APEYLLNQGTETGSDYAFGVLLLELLGKKPKVEKLAPGCQSITWAMPYLDRTKL
 PSVIDPAIKDTMDLKLHYQVAVALCVQPEPSYRPLITDVLHSLIPLVPELGGTLK
 TIKCASMDHC"
 complement (37579..37728)
 /gene="TSK6_10"
 /number=1
 complement (37729..37824)
 /gene="TSK6_10"
 /number=1
 complement (37825..37914)
 /gene="TSK6_10"
 /number=2
 complement (37915..38002)
 /gene="TSK6_10"
 /number=2
 complement (38003..38121)
 /gene="TSK6_10"
 /number=3
 complement (38122..38248)
 /gene="TSK6_10"
 /number=3
 complement (38249..38348)
 /gene="TSK6_10"
 /number=4
 complement (38349..38438)
 /gene="TSK6_10"
 /number=4
 complement (38439..38535)
 /gene="TSK6_10"
 /number=5
 complement (38536..38627)
 /gene="TSK6_10"
 /number=5
 complement (38628..38691)
 /gene="TSK6_10"
 /number=6
 complement (38692..38774)

/gene="TSK6_10"
 /number=6
 complement (38775..38907)
 /gene="TSK6_10"
 /number=7
 complement (38908..38989)
 /gene="TSK6_10"
 /number=7
 complement (38990..39169)
 /gene="TSK6_10"
 /number=8
 complement (39170..39246)
 /gene="TSK6_10"
 /number=8
 complement (39247..39386)
 /gene="TSK6_10"
 /number=9
 complement (39387..39987)
 /gene="TSK6_10"
 /number=9
 complement (39988..40123)
 /gene="TSK6_10"
 /number=10
 42133..46290
 /gene="TSK6_20"
 complement (42133..42389,42567..42618,43012..43170,
 43361..43438,43873..43944,44028..44356,44450..44600,
 44714..44773,44889..45055,45380..45452,45542..45691,
 45790..46009,46091..46290)
 /gene="TSK6_20"
 complement (42133..42389,42567..42618,43012..43170,
 43361..43438,43873..43944,44028..44356,44450..44600,
 44714..44773,44889..45055,45380..45452,45542..45691,
 45790..46009,46091..46290)
 /gene="TSK6_20"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAC03451.1"
 /db_xref="GI:9795155"
 /db_xref="SPTREMBL:Q9FY62"
 /translation="MKPAINTLVNTWAKLKSFGVMSSIEGSESTVITSPSR
 RVRELLALCPVSFLSSSSPFGSNEKFAANFVIFSVEEAGGFQDFESFVTLVSC
 DSYENVALDANNELEVEIEDEILKVLSPQMDQVVIDALSFHLKPKVTSKFADI
 SSRQLQVEIVDFRVEACNPRDMLSLCEALDAARCYHSASTSTFLLHGLSKVFIL
 IQRRHYEQLVAVPIVNLVKDLSLETDQVEDLFDKALGIASSIRDVSKINNEEA
 KVRCLLCGVYIQTATISVSIRDKAASCIPLVQLPEPLTSCGLTHGLITGNDTEKL
 MSTVAGDDDEFITSPFDISLGASLFTCAKISHEVAEANAVALGSSVVDLQNPVKFW
 QATGMLKYLISGDLLEWFEKHAIEFLDITKGTSSQCNDQDCSDYTFGIYATILQ
 AVTLLINYPADRLRKTFFALKRVLSIDIPAPHRFDVIRALVTNRSFPMALLGLV
 KDSMSKSLQDTDCAAVDTHVILVELVLRPPQGPPLPPQSDAILAALNLYRFALL
 FESRDKAICEAGKRSKVGSDILSKKNLEKAYKEWLLPLRLTLVSCSIAENLEKDHQ
 ESSLDVGLLNPIELVLYRCIELVEEKLKSH"
 complement (42133..42389)
 /gene="TSK6_20"
 /number=1
 complement (42390..42566)
 /gene="TSK6_20"
 /number=1
 complement (42567..42618)
 /gene="TSK6_20"
 /number=2
 complement (42619..43011)
 /gene="TSK6_20"
 /number=2
 complement (43012..43170)
 /gene="TSK6_20"
 /number=3
 complement (43171..43360)
 /gene="TSK6_20"
 /number=3
 complement (43361..43438)
 /gene="TSK6_20"
 /number=4

```

intron      complement(43439..43872)
            /gene="T5K6_20"
            /number=4
exon        complement(43873..43944)
            /gene="T5K6_20"
            /number=5
intron      complement(43945..44027)
            /gene="T5K6_20"
            /number=5
exon        complement(44028..44356)
            /gene="T5K6_20"
            /number=6
intron      complement(44357..44449)
            /gene="T5K6_20"
            /number=6
exon        complement(44450..44600)
            /gene="T5K6_20"
            /number=7
intron      complement(44601..44713)
            /gene="T5K6_20"
            /number=7
exon        complement(44714..44773)
            /gene="T5K6_20"
            /number=8

Query Match      46.5%; Score 220.4; DB 8; Length 96379;
Best Local Similarity 72.6%; Pred. No. 9.8e-59;
Matches 312; Conservative 0; Mismatches 91; Indels 27; Gaps 1;

QY 71 AGCGCGGAGGAAATCTGCGCGGACTACTGGAGGAGTCAAGAAACCTCACCGTTACCGTTC 130
DB 24161 AGCGCGGAGGAAATCAGCACCAACCCAGCGGAGTCAAGAAACCTCACCGTTACCGTTC 24102

QY 131 CGGACCGCTCGCTCTCTGAGATTCGTAATACCAGAGGACGACAGAGTTGTTGATCC 190
DB 24101 CAGGACCGCTCGCTCTCTGAGATTCGTAATACCAGAGGAGTCTGAATTTGTTGATCC 24042

QY 191 GTAAATCTCTTTTCAACGCTTGTTCGTAATCGCTCAAGATTCAGAGCGGATCTCA 250
DB 24041 GCAAGCTTCCCTCCAGCGCTCTGTTCCGGAATAGCTCAGGATTCAGAGTAAATTTA 23982

QY 251 GATTCAGAGCCATGCGGTGTTAGCTCTTCAAGAAGCT----- 288
DB 23981 TACTCATTAAGCGCGGATTAATAACCTTATCAATTAGTTATGTTATGGAATTAGGTATCT 23922

QY 289 -----CTCAAGCATATTTGACGATCTGAGATTCAGAGCCATGCGGTGTTAGCTCTTC 343
DB 23921 AATTAGTTAATTTTGCAGACGATCTGAGTTTCAGAGTCACGCGGTGTTAGCTCTTC 23862

QY 344 AAGAAGCTCGGAGGATATTTGGTGGTTGTTTGAAGACACAAATCTGTGCGCATTC 403
DB 23861 AAGAAGCTCGGAGGATATCTAGTTGGACTTTTCGAGGATCTAATCTCTGCGCTATTC 23802

QY 404 ATGCAAGAGGTTAGATATGCTTAAGATCTTCAATGCGCAAGGATTCGTGGAG 463
DB 23801 ACCTCAAGAGGTTAGATATGCTTAAGATCTTCAATGCGTAGAAGATTCGTGGAG 23742

QY 464 AGCGTGCTTA 473
DB 23741 AACGTGCTTA 23732

RESULT 11
ATT30N20/c
LOCUS      98473 bp      DNA      linear      PLN 06-JUL-2000
DEFINITION Arabidopsis thaliana DNA chromosome 5, BAC clone T30N20 (BSSA project).
ACCESSION AL365234
VERSION   AL365234.1 GI:8979707
KEYWORDS  Arabidopsis thaliana (thale cress)
SOURCE    Arabidopsis thaliana
ORGANISM  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 98473)
 Bevan.M., Peters.S.A., van Staveren.M., Dirkse.W., Stiekema.W., Bancroft.I., Mewes.H.W., Rudd.S., Lemcke.K. and Mayer.K.F.X. Unpublished
 2 (bases 1 to 98473)
 EU Arabidopsis sequencing project.
 Direct Submission
 Submitted (06-JUN-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de
 Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk
 Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES
 Location/Qualifiers
 1..98473
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /variety="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 1424..3476
 /gene="T30N20_10"
 complement(join(1424..1768,1854..1928,2027..2100,2220..2306,2463..2621,2720..2818,3035..3162,3397..3476))
 /gene="T30N20_10"
 complement(join(1424..1768,1854..1928,2027..2100,2220..2306,2463..2621,2720..2818,3035..3162,3397..3476))
 /gene="T30N20_10"
 /note="similarity to protein phosphatase 2C, alfalfa, PIR:T09640
 Contains Mitochondrial energy transfer proteins signature AA246-255; Protein phosphatase 2C signature AA64-72"
 /codon_start=1
 /product="protein phosphatase 2C-like protein"
 /protein_id="CAB96829.1"
 /db_xref="GI:8979708"
 /db_xref="GOA:Q9LEW5"
 /db_xref="SPTREMBL:Q9LEW5"
 /translation="MGYLDLALSYSNQPTVEAPASGGGLSQNGKFSYCYASSAGKRS SMEDPFETRIDINGEIVGLFGVDFGHCARAAEVYKRLHLSNLIITPKFISDTKSAI TDVNHDTSEILLKSNENRDAGSTASTAILVGRDLVAVNGDSPAIVSRGSKATVS RDHKPDQSDERERTENAGCFVWAGVLAVSARGLLKQYVADPEIQEIKIDDTLE FLILASGLWDVFNAAVAVMVEDEDSAKLVGEATKRGSDADNITCVVRELEK KSASSSHISSSSSEKAKEMPPGLDIALISSNEAKQVOIGSGNKPENVTNRKPDTSRST DTLTLENSVTDKV"
 complement(1424..1768)
 /gene="T30N20_10"
 /number=1
 complement(1769..1853)
 /gene="T30N20_10"
 /number=1
 complement(1854..1928)
 /gene="T30N20_10"
 /number=2
 complement(1929..2026)
 /gene="T30N20_10"
 /number=2
 complement(2027..2100)
 /gene="T30N20_10"
 /number=3
 complement(2101..2219)
 /gene="T30N20_10"
 /number=3
 complement(2220..2306)
 /gene="T30N20_10"
 /number=4
 complement(2307..2462)
 /gene="T30N20_10"

REFERENCE
 AUTHORS
 JOURNAL
 REFERENCE
 AUTHORS
 JOURNAL
 TITLE
 JOURNAL

COMMENT
 Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES
 source

exon	/number=4 complement(2463..2621) /gene="T30N20_10"	exon	/number=1 complement(9622..9793) /gene="T30N20_30"
intron	/number=5 complement(2622..2719) /gene="T30N20_10"	gene	/number=2 10959..12889 /gene="T30N20_40"
exon	/number=5 complement(2720..2818) /gene="T30N20_10"	gene	complement(join(10959..11782,11869..12292,12797..12889)) /gene="T30N20_40"
intron	/number=6 complement(2819..3034) /gene="T30N20_10"	CDS	complement(join(10959..11782,11869..12292,12797..12889)) /gene="T30N20_40"
exon	/number=6 complement(3035..3162) /gene="T30N20_10"		/note="strong similarity to nucleoid DNA-binding protein cnd41, chloroplast, common tobacco, PIR:T01996 Contains Eukaryotic and viral aspartyl proteases active site AA119-130;Prokaryotic membrane lipoprotein lipid attachment site AA155-165"
intron	/number=7 complement(3163..3396) /gene="T30N20_10"		/codon_start=1 /product="nucleoid DNA-binding protein cnd41-like protein"
exon	/number=7 complement(3397..3476) /gene="T30N20_10"		/protein_id="CAB96832.1"
gene	/number=8 6700..7608 /gene="T30N20_20"		/db_xref="GI:8979711"
CDS	6700..7608 /gene="T30N20_20"		/db_xref="GOA:O9LEW2"
	/note="similarity to predicted proteins, Arabidopsis thaliana"		/translation="MELKKERLILILKLSALSSLLHHHLVFFLPESLSLHVRHGTCS RLNKGATSPDHVEILRLDOARVNSIHSKLSKLANIDHVSSEKSTDLPAKDGSTLQSG NIIVVGLGTPKNDLSLIETGSDLTWQCQPCRYCQKPEIFNPKSTSYNVSC SSAACGSSATGNSAGSASNCIYQIQDQSFVGLAKERTLITNSDVPDGVFG CGENQGLFTGAGLLGLKLPSPQATAYNKIFSYCLPSSASYTGHLLTFSAGI SRVKTPTSTITDGTSTFYGLNVAITVGGKLPPISTVSTFGALIDSGVTITRLPP KAAVALRSSEFKAKMSKYPTSGVSLDTCFLSGKVTIPKVPFSGVAVVGLGSK GIFYVFKISQVCLAPAGNSDDSNAAIFGNVQQTLEVVYDAGGRVGFAPNGCS"
exon	/codon_start=1 /product="putative protein"	exon	complement(10959..11782) /gene="T30N20_40"
	/protein_id="CAB96830.1"	intron	/number=1 complement(11783..11868) /gene="T30N20_40"
	/db_xref="GI:8979709"	exon	/number=1 complement(11869..12292) /gene="T30N20_40"
	/db_xref="SPTREMBL:O9LEW4"	intron	/number=2 complement(12293..12796) /gene="T30N20_40"
	translation="MSPSKORHSSSTGENSKSPVRSASSAIPFWITESNNGSLRRV DPDTGDCWASPGDVPSLRSDSYLSKKQTPAGDYLSPAGMDWLKSTKLEVALR PDRVAHRAQSRGSLKSFIFAVNLQIPGKHDSVAFYFATEPPIPSGLLHRFI NGDAPFRNQRFKIVNIVGKPVVKAAGNYSACLIGKALTICNHRGPNFYFDDVDIS SSAIAITAILALGYVTSVTDIMGFLAEATEBELPILICAVRVQCMEMSSAFVYDA PPQQPLPSQPCRTLLSSAKVNHDEED"	exon	/number=2 complement(12797..12889) /gene="T30N20_40"
exon	6700..7608 /gene="T30N20_20"	gene	16050..17299
gene	8299..9793		Query Match 46.5%; Score 220.4; DB 8; Length 98473; Best Local Similarity 72.6%; Pred. No. 9.8e-59; Matches 312; Conservative 0; Mismatches 91; Indels 27; Gaps 1;
gene	complement(join(8299..9521,9622..9793))		Qy 71 AGCGCGGAGGAAATCTGCCCGACTACTCGAGGAGTCAAGAACTCACCGTTACCGTC 130
CDS	complement(join(8299..9521,9622..9793))	Db 80649 AGCGCGGAGGAAATCAGACCAACACCGGAGGAGTCAAGAACTCACCGTTACCGTC 80590	
	/note="strong similarity to nucleoid DNA-binding protein cnd41, chloroplast, common tobacco, PIR:T01996 Contains Eukaryotic and viral aspartyl proteases active site AA147-158;Prokaryotic membrane lipoprotein lipid attachment site AA42-52"	Qy 131 CCGGAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGACACAGATCTGTGATCC 190	
	/codon_start=1 /product="nucleoid DNA-binding protein cnd41-like protein"	Db 80589 CAGGAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGAGTACTGAATTTGTGATCC 80530	
	/protein_id="CAB96831.1"	Qy 191 GTAACTTCCTTTTCAACGCTTGTTCGTGAAATCGCTCAAGATTACAAGACGATCTGA 250	
	/db_xref="GI:8979710"	Db 80529 GCAGCTTCCTTCCACCGTCTCGTTCGGAATAGCTCAGGACTTCAAGTAAATTTA 80470	
	/db_xref="GOA:O9LEW3"	Qy 251 GATTCCAGAGCATCGCGGTGTAGCTCTTCAAGAAGCT-----288	
	translation="MSIMRNFLSMILMCLVCINWCPAEGAEKSDSGKVLDSYTIQVSS LFPSSSCVPSKASNTSKSRVVMHGACSHLSSDARVHDEILIRDQARVEIYSK LSKNSANEVSPAKETLPAKSGITIGSGNYITVIGTPRHDSLVDFTGSDLTWTC EPLCGSCYQKPEFNSSSYQYVSCSPMCDABSCASNCVSIIVYGDKSFYTG FLAKEFTLTNSDVLVDVYFCGNNQGLFDGVLGLGPKLSLPAQTTTNNIF SYCLPSTNSSTGHLTPGSAGISESVKFTPLSSPSAPNYGIDIIIGISVGDKEAINTP NSFTFGAIDSGGVTFILPKKYABLRVSRVFERKMSYKSTSGVGLDTCDFDTGLDI VYTIAPFSFVVELDGSILPIKISQVCLAFAGNDLPAIFGNVQQTLLDVIY DVAGRVGFAPNGC"	Db 80469 TACTCATTAGCGCGGATTAATACTTATCAATAGGTTATGTTATGAATTAGGTACT 80410	
	/db_xref="SPTREMBL:O9LEW3"	Qy 289 -----GCTGAAGCATATTTCACGATCTGAGATTCAGAGCCATCGCGTGTAGCTTTC 343	
	complement(8299..9521)	Db 80409 AATTAGGTTAAATTTTCAGACCGATCTGAGGTTTCAGAGTCAACCGGTGTAGCTTTC 80350	
exon	/gene="T30N20_30"	Qy 344 AAGAAGCTCGGAGGCATATTTCGTGGGTTTGTGGAAGACACAAATCTGTGTCCCATTC 403	
intron	/number=1 complement(9522..9621) /gene="T30N20_30"	Db 80349 AAGAGCGGCTGAGGCATACCTAGTTGACTTTTCGAGGATACATCTCTGCGCTATTC 80290	


```

/gene="histone H3.3"
/number=1
2568. .2861
/gene="histone H3.3"
/number=2
join(2580. .2861,2928. .3056)
/gene="histone H3.3"
/codon_start=1
/protein_id="CAA57081.1"
/db_xref="GI:963026"
/db_xref="FLYBASE:FBgn0014722"
/db_xref="GOA:P06351"
/db_xref="SWISS-PROT:P06351"
/translation="WATKOPARKSTGGKAPRKOLATKAARKSPSTGGVKKPHRYRP
GTVALREIRRYKQTELLIRKLPFORLVREIAQDFKDLRFQSAIGALQBASEAYLV
GLFEDTLNLCALHAKRVIMPDKDIQLARRIRGERA"
intron
2862. .2927
/gene="histone H3.3"
/number=2
2928. >3056
/gene="histone H3.3"
/number=3

ORIGIN
Query Match 45.3%; Score 214.8; DB 3; Length 4534;
Best Local Similarity 68.5%; Pred. No. 3.8e-57;
Matches 326; Conservative 0; Mismatches 147; Indels 3; Gaps 2;

Qy 1 ATGGCTGTGACAGCAAACTGCAAGAAATCACACGAGGAGAAAGCTCCGAGACTCTG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2580 ATGGCAGCTACCAAGCAGACGCCCGTAAATCAACTGGAGGCAAGGCTCTCGAAGCAA 2639
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 CTCGCTACCAAGCGGCGAGGAAATCTCGCGGACTACTGAGGAGTCAAGAACTTCAAC 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2640 CTGGCCACTAAGGAGCGCCGTAATCGGCCCGCTCCACAGGTGGTGTAAAGAGCCCAT 2699
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 CGTTACCGTCCGGAACCGTCTCTCGTGAGATTCTGTAATACCAAGAGCACAGAG 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2700 CGTTATCGTCCCGTACTGTGCTTTGGTGAGATCCGTCTACCAAGAAATCGACGAG 2759
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 TTGTGATCGGTAAACTCTCTTTTCAACAGTCTGTGTGTAATCGCTCAAGATTACAAG 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2760 TTGCTCATCCGCAAACTCGCTTTCAACAGGTGGTGGTGAATCGCAAGATTTCAG 2819
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 ACGGATCTGAGATTCACAGGCAATCGCGTGTGTAGCTCTTCAAGAGCTGC--TGAAGCAT 298
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2820 ACCGATTGGTTCCTTCAGTCCGAGCTATCGGTGCTTGCAGGTAAGCACCAGAAACACT 2879
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
299 ATTTGACGGATCTGAGATTCACAGGCCATGCGGT-GTTAGCTCTTCAAGAGCTCGGAG 357
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2880 ATTTAAACCAATTGAATTTCAATAATACGAAATAACTAATGTTGCAGGAAGCCTCCGAG 2939
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
358 GCATATTTGGTGGTGTGTTTGAAGACACAATCTGTGTGCATCTATCAAGAGGGTT 417
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2940 GCCTATTTGGTGGCTTATTTGAAGACACCAACTTTGTGCGCATTTACGCGCAAGCGTCTC 2999
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
418 ACGATAATGCGCTAAAGATGTTCAATTGGCAAGAGATTTCGTGGAGAGCGTGCCTTA 473
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3000 ACCATTATGCCAAGACATTCAGTTGGCGCTCGCATTCGCGGAGCGGTGCTTA 3055

RESULT 14
AC130200
LOCUS AC130200 119578 bp DNA linear HTG 11-SEP-2003
DEFINITION Medicago truncatula clone mthl-64n13, WORKING DRAFT SEQUENCE.
ACCESSION AC130200
VERSION AC130200.28 GI:33147116
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

```

```

Medicago.
1 (bases 1 to 119578)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Medicago truncatula BAC Clone mthl-64n13
Unpublished
2 (bases 1 to 119578)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (08-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 119578)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (11-SEP-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Jul 23, 2003 this sequence version replaced gi:32567827.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 119578: contig of 119578 bp in length.
FEATURES
source
1. 119578
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mthl-64n13"
/clone_lib="Medicago truncatula BAC library H1"
/note="This is one of two clones in the same well from
mthl-64n13"

ORIGIN
Query Match 44.3%; Score 210; DB 2; Length 119578;
Best Local Similarity 69.8%; Pred. No. 2.2e-55;
Matches 331; Conservative 0; Mismatches 80; Indels 63; Gaps 1;

Qy 1 ATGGCTGTGACAGCAAACTGCAAGAAATCACACGAGGAGAAAGCTCCGAGACTCTG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96853 ATGGCAGCTACCAAAACACAGACGCAATTCACCGCGGAGAAAGCCCAACAA 96912
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 CTCGCTACCAAGCGGCGGAGAAATCTCGCGGACTACTGAGGAGTCAAGAAACCTTCAAC 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96913 CTGCGCACCAAGACCGCACGGAATCAGTCCGCGCAGCGGAGGAGTGAAGAGCCACAC 96972
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 CGTTACCGTCCGGAACCGTCTCTCGTGAGATTCTGTAATACCAAGAGAGACAGAG 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96973 CGTTCCGTCGCGGACCGGTGGCGCTTCGTGAATCCGGAATACCAAGAGACACTGAG 97032
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 TTGTTGATCGTAACTTCCTTTTCAACGTCTTCTGCGTGAATCGCTCAAGATTACAAG 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
97033 CTTCTGATAGAAACCTTCATTCCAACGCTTGTTCGTGAATTCGCTCAGGATTTCAA 97092
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 ACGGATCTGAGATTCACAGAGCCATCGCGTGTGTAGCTCTTCAAGAGAGTGTGTAAGCATAT 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
97093 ACCGATCTCAGATTCACAGAGCAGTGTGTTCTGCTCTTCAAGAGAGCGGTGAA----- 97146
Qy ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 TTGACGATCTGAGATTCACAGAGCCATCGCGTGTGTAGCTCTTCAAGAGAGTTCGCGAGGCA 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 97147 -----GCT 97149
QY 361 TATTGGTGGTTGTTGAAGACACAAATCTGTGTCCTATTCATGCAAGAGGGTTACG 420
Db 97150 TATCTTGTGTTGTTGTTGAAGATACAAATCTTTGCGCTATTTCATGCAAGAGTTACT 97209
QY 421 ATAATGCTAAAGATGTTCAATTGGCAGAAGGATTGCTGAGAGCGTGTCTTAG 474
Db 97210 ATTATGCTAAGATATTCATCTGCAAGAGAAATTCGTGTGAAGGGCTTAG 97263

RESULT 15
LOCUS AY078027 411 bp mRNA linear PLN 18-FEB-2002
DEFINITION Arabidopsis thaliana AT4g40030/T5J17_200 mRNA, complete cds.
ACCESSION AY078027
VERSION AY078027.1 GI:18700235
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsie.
1 (bases 1 to 411)
Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Banh,J., Bowser,L.,
Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K.,
Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A.
and Ecker,J.R.
Arabidopsis ORF clones
Unpublished
2 (bases 1 to 411)
Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Banh,J., Bowser,L.,
Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y.,
Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J.,
Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M., Nguyen,M.,
Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T., Satou,M., Seki,M.,
Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K.,
Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A.
and Ecker,J.R.
Direct Submission
Submitted (04-FEB-2002) Salk Institute Genomic Analysis Laboratory
(SiGnAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGSC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Cheuk,R., Chen,H.,
Kim,C.J., Meyers,M.C., Shinn,P., Banh,J., Bowser,L., Chan,M.M.,
Chang,E., Dale,J.M., Deng,J.M., Goldsmith,A.D., Jones,T.,
Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M., Nguyen,M.,
Onodera,C.S., Palm,C.J., Quach,H.L., Southwick,A., Tang,C.C.,
Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu,G., Davis,R.W.,
Theologis,A., and Ecker,J.R.

Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
contributed equally to this work as PIs.
Location/Qualifiers
1. .411
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="4"
/clones="U12616"

FEATURES
source

```

```

/ecotype="Columbia"
/notes="This clone is in puni 51"
1. .411
/notes="histone H3.3"
/codon_start=1
/product="AT4g40030/T5J17_200"
/protein_id="AAL77728.1" _200"
/db_xref="GI:18700235"
/translation="MARTKQTARKSTGGKAPRKQLATKAARKSAPTTGGVKKPHRRYP
GTVALREIRKYQKSTELLIRKUPFPQRLVEIAQDFKTLRFQSHVALQEAEEAIVL
GLFEDTNLCIAHAKRVTIMEKDIQLARRINGERA"

ORIGIN
Query Match 44.0%; Score 208.6; DB 8; Length 411;
Best Local Similarity 79.4%; Pred. No. 2.6e-55; Indels 0; Gaps 0;
Matches 247; Conservative 0; Mismatches 64;

QY 1 ATGGCTCGTACGAAGCAAACTGCAAGAAATCACAGGAGGAAAAGCTCCGAGGACTCTG 60
Db 1 ATGGCTCGTACGAAGCAAAACCGCTCGTAAGTCCACCGGAGGTAAGCTCCAAGGAACAA 60
QY 61 CTGCTACCAAGCGGGGAGGAAATCTGCGCCGACTACTGCGAGGAGTCAAGAAACCTCAC 120
Db 61 CTTGCTACTAAGGCTGCTCGTAAATCTGCACCAACTACTGCTGGAGTCAAGAAACCAAT 120
QY 121 CGTTACCGTCCCGGAACCGTGGCTCTTCGTGAGATTCGTAATACCAAGAGACACAGAG 180
Db 121 CGTTACCGTCCCGGAACCGTGGCTCTTCGTGAGATTCGTAATACCAAGAGAGTACTGAG 180
QY 181 TTGTTGATCGTAAACTTCCTTTTCAACGCTCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 240
Db 181 TTGTTGATCGTAAACTTCCTTTTCAACGCTCTTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 240
QY 241 ACGGATCTGAGATTCAGAGCCATGCGGTGTTAGTCTCTTCAAGAGCTGCTGAAGATAT 300
Db 241 ACTGATTTGGTTTCCAGAGCCATGCGGTGTTAGTCTCTTCAAGAGCTGCTGAAGATAT 300
QY 301 TTGACGGGATCT 311
Db 301 CTTGTTGGTCT 311

Search completed: June 28, 2004, 16:23:04
Job time : 2085.37 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2004, 14:35:24 ; Search time 1643.18 Seconds
(without alignments)
8614.203 Million cell updates/sec

Title: AAC50492

Perfect score: 474

Sequence: 1 ATGGCTGCTACGAGCAAAAC.....TTGCTGGAGAGCGTGCTTAG 474

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estov.*
5: em_estpl.*
6: em_estro.*
7: em_estro.*
8: em_estro.*
9: gb_est1.*
10: gb_est2.*
11: gb_est3.*
12: gb_est4.*
13: gb_est5.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vrt.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrt.*
28: gb_gss1.*
29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	259.6	54.8	668	14	CD830068
2	256.4	54.1	572	14	CD820974
3	256.4	54.1	604	14	CD818139
4	256.4	54.1	622	14	CD834370

5	256.4	54.1	630	14	CD838405
6	256.4	54.1	635	14	CD842690
7	256.4	54.1	637	14	CD834027
8	256.4	54.1	639	14	CD838656
9	256.4	54.1	676	14	CD823357
10	255.4	53.9	527	9	AA650659
11	237.8	50.2	545	13	BQ791478
12	233	49.2	545	9	AV552968
13	233	49.2	603	14	CD830234
14	233	49.2	611	9	AI998977
15	233	49.2	643	14	CD822585
16	233	49.2	664	9	AV824078
17	232.6	49.1	414	13	BQ791662
18	231.4	48.8	593	14	CD829010
19	231.4	48.8	594	14	CD829250
20	231.4	48.8	599	14	CD830791
21	231.4	48.8	633	14	CD824194
22	231.4	48.8	634	14	CD819611
23	231.4	48.8	669	14	CD818523
24	230.2	48.6	606	9	AA713310
25	229.8	48.5	599	14	CD824297
26	229.8	48.5	605	14	CD819641
27	228.8	48.3	615	14	CD820205
28	228.8	48.3	689	14	CD839282
29	224.2	47.3	547	14	CD821103
30	223.6	47.2	453	9	AA713299
31	222.8	47.0	447	10	BE317214
32	222.8	47.0	501	10	BF643573
33	222.8	47.0	505	12	BI263311
34	222.8	47.0	571	10	AW574054
35	222.8	47.0	628	12	BG454253
36	222.8	47.0	630	12	BI263220
37	222.8	47.0	635	9	AJ496962
38	222.8	47.0	638	14	CA319769
39	222.8	47.0	656	12	BG453911
40	222.8	47.0	666	13	BQ165501
41	222.8	47.0	667	14	CF069311
42	222.8	47.0	670	12	BI271678
43	222.8	47.0	676	10	BE325005
44	222.8	47.0	678	12	BG455610
45	222.8	47.0	693	12	BI272111

ALIGNMENTS

RESULT 1
CD830068

LOCUS

DEFINITION

BN40.044E03F011221 BN40 Brassica napus cDNA clone BN40044E03, mRNA

sequence.

CD830068

CD830068.1 GI:32512008

EST.

Brassica napus (rape)

Brassica napus

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 668)

Genoplatte.

Genoplatte, a major partnership french program in plant genomics

Unpublished (2003)

Contact: Genoplatte

Genoplatte

93, rue Henri Rochefort 91025 EVRY CEDEX France

Tel: 33 1 69 47 54 00

Fax: 33 1 69 47 54 10

This sequence has been generated in the framework of the french

plant genomics programme 'Genoplatte' (<http://www.genoplatte.com>

and <http://genoplatte-info.infobiogen.fr>).

Location/Qualifiers

source

1.668

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="Jet neuf"
/db_xref="taxon:3708"
/clone="BN40044E03"
/tissue_type="seed"
/clone_lib="BN40"

ORIGIN

Query Match 54.8%; Score 259.6; DB 14; Length 669;
Best Local Similarity 76.4%; Pred. No. 4.5e-66;
Matches 362; Conservative 0; Mismatches 49; Indels 63; Gaps 1;

QY 1 ATGGCTCTACCAAGCAAACTCGAGAAATCACCGAGGAAAGCTCCGAGGACTCTG 60
DB 122 ATGGCTCTACCAAGCAGACTCGAGAAATCACCGAGGAAAGCTCCGAGGAAAGCA 181
QY 61 CTCGCTTACCAAGCGCGGAGGAAATCTGCGCGACTACTGAGGAGTCAAGAAACCTCAC 120
DB 182 CTCGCCACCAAGCTCGAGGAAATCAGACCGACCCCGAGGAGTCAAGAGCTCTAC 241
QY 121 CGTTACCGTCCCGGAAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGCACAGAG 180
DB 242 CGTTACCGTCCCGGAAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGCACAGAG 301
QY 181 TTGTTGATCGTAAACCTCTCTTCAACGCTTGTTCGTGAAATCGTCAAGATTACAG 240
DB 302 TTATGATCCGTAACCTCTCTTCCAGCGTCTGTTCGTGAAATCGCCAGGATTTCAA- 360
QY 241 ACGGATCTGAGATTCAGAGCCATCGGCTGTAGTCTTCAAGAGCTGCTGAAGCATAT 300
DB 361 ----- 360
QY 301 TTGACGATCTGAGATTCAGAGCCATCGGCTGTAGTCTTCAAGAGCTGCGGAGGCA 360
DB 361 --GACGATCTGAGATTCAGAGCCATCGGCTGTAGTCTTCAAGAGCTGCGGAGGCG 418
QY 361 TATTGTTGGTGGTGTGTTTGAAGACACAAATCTGTGTCATTCATCAAGAGGTTAGC 420
DB 419 TACTTGGTGGTGTGTTTGAAGACACAAATCTGTGTCATTCATCAAGAGGTTAGC 478
QY 421 ATAATGCTTAAAGATTTCAATGCGCAAGAGATTCGTGGAGAGCGTCTTAG 474
DB 479 ATCATGCCCAAGACATTCATTTGGCTAGGCGCATTCGTGGAGAGCGTCTTAG 532

RESULT 2

CD820974
LOCUS BN25.040G21F011026 BN25 Brassica napus cDNA clone BN25040G21, mRNA
DEFINITION sequence.
ACCESSION CD820974
VERSION CD820974.1 GI:32502914
KEYWORDS EST.
SOURCE Brassica napus (rape)
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 572)
Genoplante.
AUTHORS Genoplante, a major partnership french program in plant genomics
TITLE Unpublished (2003)
JOURNAL
COMMENT Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (<http://www.genoplante.com>)
and <http://genoplante-info.infobiogen.fr>.
FEATURES Location/Qualifiers
source 1..572

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="Jet neuf"
/db_xref="taxon:3708"
/clone="BN25040G21"
/tissue_type="seed"
/clone_lib="BN25"

ORIGIN

Query Match 54.1%; Score 256.4; DB 14; Length 572;
Best Local Similarity 75.9%; Pred. No. 3.8e-65;
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;

QY 1 ATGGCTCTACCAAGCAAACTCGAGAAATCACCGAGGAAAGCTCCGAGGACTCTG 60
DB 56 ATGGCTCTACCAAGCAGACTCGAGAAATCACCGAGGAAAGCTCCGAGGAAAGCA 115
QY 61 CTCGCTTACCAAGCGCGGAGGAAATCTGCGCGACTACTGAGGAGTCAAGAAACCTCAC 120
DB 116 CTCGCCACCAAGCTCGAGGAAATCAGACCGACCCCGAGGAGTCAAGAGCTCTAC 175
QY 121 CGTTACCGTCCCGGAAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGCACAGAG 180
DB 176 CGTTACCGTCCCGGAAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGCACAGAG 235
QY 181 TTGTTGATCGTAAACCTCTCTTCAACGCTTGTTCGTGAAATCGTCAAGATTACAG 240
DB 236 TTGTTGATCGGCAAGCTTCCGTTTCAGCGTCTGTTCGTGAAATCGCCAGGATTTCAA- 294
QY 241 ACGGATCTGAGATTCAGAGCCATCGGCTGTAGTCTTCAAGAGCTGCTGAAGCATAT 300
DB 295 ----- 294
QY 301 TTGACGATCTGAGATTCAGAGCCATCGGCTGTAGTCTTCAAGAGCTGCGGAGGCA 360
DB 295 --GACGATCTGAGATTCAGAGCCATCGGCTGTAGTCTTCAAGAGCTGCGGAGGCG 352
QY 361 TATTGTTGGTGGTGTGTTTGAAGACACAAATCTGTGTCATTCATCAAGAGGTTAGC 420
DB 353 TACTTGGTGGTGTGTTTGAAGACACAAATCTGTGTCATTCATCAAGAGGTTAGC 412
QY 421 ATAATGCTTAAAGATTTCAATGCGCAAGAGATTCGTGGAGAGCGTCTTAG 474
DB 413 ATCATGCCCAAGACATTCATTTGGCTAGGCGCATTCGTGGAGAGCGTCTTAG 466

RESULT 3

CD818139
LOCUS BN20.044G19F011127 BN20 Brassica napus cDNA clone BN20044G19, mRNA
DEFINITION sequence.
ACCESSION CD818139
VERSION CD818139.1 GI:32500079
KEYWORDS EST.
SOURCE Brassica napus (rape)
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 604)
Genoplante.
AUTHORS Genoplante, a major partnership french program in plant genomics
TITLE Unpublished (2003)
JOURNAL
COMMENT Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (<http://www.genoplante.com>)
and <http://genoplante-info.infobiogen.fr>.
FEATURES Location/Qualifiers
source 1..604

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="Jet neuf"
/db_xref="taxon:3708"
/clone="BN20044G19"
/tissue_type="seed"
/clone_lib="BN20"

ORIGIN

```

Query Match      54.1%; Score 256.4; DB 14; Length 604;
Best Local Similarity 75.9%; Pred. No. 3.8e-65;
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;

Qy 1 ATGGCTGTACGACCAACTGCAAGAAATCAACGAGGAGAAAGTCCGAGGACTCTG 60
Db 62 ATGGCTGTACGACGAGACTGCTAGGAAATCCACCGGAGGAAAGTCCGAGGAGCAA 121
Qy 61 CTGCTACCAAGGCGGAGGAAATCTGCGCGGACTACTGAGGAGTCAAGAACTCTCAC 120
Db 122 CTGCGCCACCAAGGCGGAGGAAATCGGCGCGGACCGAGGAGTCAAGAGCTCTCAC 181
Qy 121 GGTTCACGTCCTCCGAAACCGTCGCTCTTCGTGAGATTCGTAATACCAAGAGACAGAG 180
Db 182 CGTTACCGTCCCGAAACCGTCGCTCTTCGTGAGATTCGTAAGTACCAAGAGACTGAG 241
Qy 181 TTGTTGATCCGTAACCTCTTTCAACGCTCTTCTGCGGAAATCGCTCAAGATTACAG 240
Db 242 TTGTTGATCCGTAAGCTTCCGTTTCAGCGTCTGTTCTGAGATTCGCCAGGATTTCAA- 300
Qy 241 ACGGATCTGAGATTCAGAGCCATGCGGTGTTAGCTCTTCAAGAAGCTGCTGAAGCATAT 300
Db 301 ----- 300

Qy 301 TTGACGATCTGAGATTCAGAGCCATGCGGTGTTAGCTCTTCAAGAAGCTGCGAGGCA 360
Db 301 --GACAGACCTGAGGTTCCAGAGCCACGACGTTGCGGCTTCAGGAGCTGCGAGGCT 358
Qy 361 TATTGTTGGGTTGTTTGAAGACACAAATCTGTGCGCATTCATGCAAGAGGTTACG 420
Db 359 TATCTGTTGGGTTGTTTGAAGACACAAATCTGTGCGCATTCATGCGGATTCAGCCAGAGAGTACC 418
Qy 421 ATAATGCTTAAAGATGTTCAATTGSCAAGAGGATTCGTGAGAGCGTGCTTAG 474
Db 419 ATCATGCTTAAAGATGTTCAATTGCTAGGCGTATTCGTGAGAGCGTGCTTAG 472

```

```

RESULT 4
CD834370
LOCUS      BN45.041116F011207 BN45 Brassica napus cDNA clone BN45041116, mRNA
DEFINITION
sequence.
ACCESSION  CD834370.1 GI:32516310
VERSION     CD834370.1
KEYWORDS   EST.
SOURCE     Brassica napus (rape)
ORGANISM   Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE  1 (bases 1 to 622)
AUTHORS   Genoplante.
TITLE     Genoplante, a major partnership french program in plant genomics
JOURNAL   Unpublished (2003)
COMMENT   Contact: Genoplante
          Genoplante
          93, rue Henri Rochefort 91025 EVRY CEDEX France
          Tel: 33 1 69 47 54 00
          Fax: 33 1 69 47 54 10
          This sequence has been generated in the framework of the french
          plant genomics programme 'Genoplante' (http://www.genoplante.com
          and http://genoplante-info.inbio.gen.fr).
          Location/Qualifiers
          source          1..622

```

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="Jet neuf"
/db_xref="taxon:3708"
/clone="BN45041116"
/tissue_type="seed"
/clone_lib="BN45"

ORIGIN

```

Query Match      54.1%; Score 256.4; DB 14; Length 622;
Best Local Similarity 75.9%; Pred. No. 3.9e-65;
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;

Qy 1 ATGGCTGTACGACCAACTGCAAGAAATCAACGAGGAGAAAGTCCGAGGACTCTG 60
Db 63 ATGGCTGTACGACGAGACTGCTAGGAAATCCACCGGAGGAAAGTCCGAGGAGCAA 122
Qy 61 CTGCTACCAAGGCGGAGGAAATCTGCGCGGACTACTGAGGAGTCAAGAACTCTCAC 120
Db 123 CTGCGCCACCAAGGCGGAGGAAATCGGCGCGGACCGAGGAGTCAAGAGCTCTCAC 182
Qy 121 GGTTCACGTCCTCCGAAACCGTCGCTCTTCGTGAGATTCGTAATACCAAGAGACAGAG 180
Db 183 CGTTACCGTCCCGAAACCGTCGCTCTTCGTGAGATTCGTAAGTACCAAGAGACTGAG 242
Qy 181 TTGTTGATCCGTAACCTCTTTCAACGCTCTTCTGCGGAAATCGCTCAAGATTACAG 240
Db 243 TTGTTGATCCGTAAGCTTCCGTTTCAGCGTCTGTTCTGAGATTCGCCAGGATTTCAA- 301
Qy 241 ACGGATCTGAGATTCAGAGCCATGCGGTGTTAGCTCTTCAAGAAGCTGCTGAAGCATAT 300
Db 302 ----- 301

Qy 301 TTGACGATCTGAGATTCAGAGCCATGCGGTGTTAGCTCTTCAAGAAGCTGCGAGGCA 360
Db 302 --GACAGACCTGAGGTTCCAGAGCCACGACGTTGCGGCTTCAGGAGCTGCGAGGCT 359
Qy 361 TATTGTTGGGTTGTTTGAAGACACAAATCTGTGCGCATTCATGCAAGAGGTTACG 420
Db 360 TATCTGTTGGGTTGTTTGAAGACACAAATCTGTGCGCATTCATGCGGATTCAGCCAGAGAGTACC 419
Qy 421 ATAATGCTTAAAGATGTTCAATTGSCAAGAGGATTCGTGAGAGCGTGCTTAG 474
Db 420 ATCATGCTTAAAGATGTTCAATTGCTAGGCGTATTCGTGAGAGCGTGCTTAG 473

```

```

RESULT 5
CD838405
LOCUS      RFO2.111B20F010523 RFO2 Brassica napus cDNA clone RFO2111B20, mRNA
DEFINITION
sequence.
ACCESSION  CD838405.1 GI:32520345
VERSION     CD838405.1
KEYWORDS   EST.
SOURCE     Brassica napus (rape)
ORGANISM   Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE  1 (bases 1 to 630)
AUTHORS   Genoplante.
TITLE     Genoplante, a major partnership french program in plant genomics
JOURNAL   Unpublished (2003)
COMMENT   Contact: Genoplante
          Genoplante
          93, rue Henri Rochefort 91025 EVRY CEDEX France
          Tel: 33 1 69 47 54 00
          Fax: 33 1 69 47 54 10
          This sequence has been generated in the framework of the french
          plant genomics programme 'Genoplante' (http://www.genoplante.com
          and http://genoplante-info.inbio.gen.fr).
          Location/Qualifiers
          source          1..630

```

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="samourai (restored line)"
/db_xref="taxon:3708"
/clone="RFO2111B20"
/tissue_type="anthers"
/clone_lib="RFO2"

ORIGIN

Query Match 54.1%; Score 256.4; DB 14; Length 630;
Best Local Similarity 75.9%; Pred. No. 3.9e-65;
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;

QY 1 ATGGCTCGTACGAGCAAACTGCAAGAAATCACAAGGAGAAAGCTCCGAGGACTCTG 60
DB 63 ATGGCTCGTACGAGCAGACTCTAGGAAATCACCAGGAGAAAGCTCCGAGGAGCAA 122
QY 61 CTCGCTACCAAGCGCGAGGAAATCTGCCCGACTACTGGAGGAGTCAAGAAACCTCAC 120
DB 123 CTCGCCACCAAGCGCGAGGAAATCGGCCCGACCAACGGAGAGTCAAGAGCTCAC 182
QY 121 CGTTACCGTCCCGAACCCTGCTCTTCGAGATTCGTAATACCAAGAGACACAGAG 180
DB 183 CGTTACCGTCCCGAACCCTGCTCTTCGAGATTCGTAATACCAAGAGACACAGAG 242
QY 181 TTGTTGATCGTAACTCTCTTTCAACGCTCTGTTGTAATCGTCAAGATTACAG 240
DB 243 TTGTTGATCGTAACTCTCTTTCAACGCTCTGTTGTAATCGTCAAGATTACAG 301
QY 241 ACGGATCTGAGATTCAGAGCCATCGCGTGTAGTCTTCAAGAGCTGCTGAAGCATAT 300
DB 302 ----- 301
QY 301 TTGACGATCTGAGATTCAGAGCCATCGCGTGTAGTCTTCAAGAGCTGCGAGGCA 360
DB 302 --GACAGACTGAGGTTCCAGAGCCAGCAGTGTGGCGCTTCAGGAAGTGGGAGGCT 359
QY 361 TATTTCGTGGTGTGTTGTTGAAGACACAAATCTCTGTGCCATTCATCAAGAGGGTTAG 420
DB 360 TATCTGTTGGTGTGTTGTTGAAGACACAAATCTCTGTGCCATTCATCAAGAGGGTTAG 419
QY 421 ATAATGCTTAAAGATTTCAATTTGCAAGAGGATTCGTGGAGAGCGTCTTAG 474
DB 420 ATCATGCTTAAAGATTTCAATTTGCTAGCGGTATTCGTGGAGAGCGTCTTAG 473

RESULT 6

CD842690
LOCUS RFO2.128J10F010914 RFO2 Brassica napus cDNA clone RFO2128J10, mRNA
DEFINITION sequence.
ACCESSION CD842690
VERSION CD842690.1 GI:32524630
KEYWORDS EST.
SOURCE Brassica napus (rape)
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 635)
AUTHORS Genoplante.
TITLE Genoplante, a major partnership french program in plant genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (<http://www.genoplante.com>
and <http://genoplante-info.inbioogen.fr>).
FEATURES
Location/Qualifiers
source 1. .635

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="samourai (restored line)"
/db_xref="taxon:3708"
/clone="RFO2128J10"
/tissue_type="anthers"
/clone_lib="RFO2"

ORIGIN

Query Match 54.1%; Score 256.4; DB 14; Length 635;
Best Local Similarity 75.9%; Pred. No. 3.9e-65;
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;

QY 1 ATGGCTCGTACGAGCAAACTGCAAGAAATCACAAGGAGAAAGCTCCGAGGACTCTG 60
DB 63 ATGGCTCGTACGAGCAGACTCTAGGAAATCACCAGGAGAAAGCTCCGAGGAGCAA 122
QY 61 CTCGCTACCAAGCGCGAGGAAATCTGCCCGACTACTGGAGGAGTCAAGAAACCTCAC 120
DB 123 CTCGCCACCAAGCGCGAGGAAATCGGCCCGACCAACGGAGAGTCAAGAGCTCAC 182
QY 121 CGTTACCGTCCCGAACCCTGCTCTTCGAGATTCGTAATACCAAGAGACACAGAG 180
DB 183 CGTTACCGTCCCGAACCCTGCTCTTCGAGATTCGTAATACCAAGAGACACAGAG 242
QY 181 TTGTTGATCGTAACTCTCTTTCAACGCTCTGTTGTAATCGTCAAGATTACAG 240
DB 243 TTGTTGATCGTAACTCTCTTTCAACGCTCTGTTGTAATCGTCAAGATTACAG 301
QY 241 ACGGATCTGAGATTCAGAGCCATCGCGTGTAGTCTTCAAGAGCTGCTGAAGCATAT 300
DB 302 ----- 301
QY 301 TTGACGATCTGAGATTCAGAGCCATCGCGTGTAGTCTTCAAGAGCTGCGAGGCA 360
DB 302 --GACAGACTGAGGTTCCAGAGCCAGCAGTGTGGCGCTTCAGGAAGTGGGAGGCT 359
QY 361 TATTTCGTGGTGTGTTGTTGAAGACACAAATCTCTGTGCCATTCATCAAGAGGGTTAG 420
DB 360 TATCTGTTGGTGTGTTGTTGAAGACACAAATCTCTGTGCCATTCATCAAGAGGGTTAG 419
QY 421 ATAATGCTTAAAGATTTCAATTTGCAAGAGGATTCGTGGAGAGCGTCTTAG 474
DB 420 ATCATGCTTAAAGATTTCAATTTGCTAGCGGTATTCGTGGAGAGCGTCTTAG 473

RESULT 7

CD834027
LOCUS BN45.040F17F011019 BN45 Brassica napus cDNA clone BN45040F17, mRNA
DEFINITION sequence.
ACCESSION CD834027
VERSION CD834027.1 GI:32515967
KEYWORDS EST.
SOURCE Brassica napus (rape)
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 637)
AUTHORS Genoplante.
TITLE Genoplante, a major partnership french program in plant genomics
JOURNAL Unpublished (2003)
COMMENT Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (<http://www.genoplante.com>
and <http://genoplante-info.inbioogen.fr>).
FEATURES
Location/Qualifiers
source 1. .637

```

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="Jet neuf"
/db_xref="taxon:3708"
/clone="BN45040R17"
/tissue_type="seed"
/clone_lib="BN45"

```

ORIGIN

```

Query Match      54.1%; Score 256.4; DB 14; Length 637;
Best Local Similarity 75.9%; Pred. No. 3.9e-65;
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;

QY 1 ATGGCTCGTACGAAGCAAACTGCAAGAAATCACACGAGGAAAGCTCCGAGGACTCTG 60
Db 63 ATGGCTCGTACGAAGCAAGCTGTAGGAAATCCACCGGAGGAAAGCTCCGAGGACAA 122
QY 61 CTGCTACCAAGCGGCGAGGAAATCTGCGCGACTACTGAGGAGTCAAGAACTCTCAC 120
Db 123 CTGCTACCAAGCGGCGAGGAAATCTGCGCGACTACTGAGGAGTCAAGAACTCTCAC 182
QY 121 CGTTACCGTCCCGAAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGACACAGAG 180
Db 183 CGTTACCGTCCCGAAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGACACAGAG 242
QY 181 TTGTTGATCCGTAAACTCTCTTTTCAACGCTCTGTTGTAATCGCTCAAGATTACAAG 240
Db 243 TTGTTGATCCGTAAAGCTCTCGTTTTCAGCGTCTGTTGTAATCGCTCAAGATTACAAG 301
QY 241 ACGGATCTGAGATTCAGAGCCATCGGTGTTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
Db 302 ----- 301
QY 301 TTGACGATCTGAGATTCAGAGCCATCGGTGTTAGCTCTTCAAGAGCTGCGAGGCA 360
Db 302 --GACAGACCTGAGTTCAGAGCCACGACGTTGTCGCTTTCAGGAGCTGCGAGGCT 359
QY 361 TATTGTTGGTGTGTTGTTGAAGACACAAATCTGTGTCCTTCAATGCAAGAGGTTACG 420
Db 360 TATTGTTGGTGTGTTGTTGAAGACACAAATCTGTGTCCTTCAATGCAAGAGGTTACG 419
QY 421 ATAATGCTTAAAGATGTTCAATTCGCAAGAGGATTCGTGAGAGCGTCTTAG 474
Db 420 ATCATGCTTAAAGATGTTCAATTCGCTAGGCTATTCTGTTGAGAGCGTCTTAG 473

```

```

RESULT 8
CD838656      639 bp      mRNA      linear      EST 10-JUL-2003
LOCUS         RF02.112B20F010523 RF02 Brassica napus cDNA clone RF02112B20, mRNA
DEFINITION   sequence.
ACCESSION    CD838656
VERSION      CD838656.1 GI:32520596
KEYWORDS     EST.
SOURCE       Brassica napus (rape)
ORGANISM     Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE    1 (bases 1 to 639)
AUTHORS      Genoplante.
TITLE        Genoplante, a major partnership french program in plant genomics
JOURNAL      Unpublished (2003)
COMMENT      Contact: Genoplante
              Genoplante
              93, rue Henri Rochefort 91025 EVRY CEDEX France
              Tel: 33 1 69 47 54 00
              Fax: 33 1 69 47 54 10
              This sequence has been generated in the framework of the french
              plant genomics programme 'Genoplante' (http://www.genoplante.com
              and http://genoplante-info.inbio.gen.fr).
              Location/Qualifiers
              source          1. .639

```

```

/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="samourai (restored line)"
/db_xref="taxon:3708"
/clone="RF02112B20"
/tissue_type="anthers"
/clone_lib="RF02"

```

ORIGIN

```

Query Match      54.1%; Score 256.4; DB 14; Length 639;
Best Local Similarity 75.9%; Pred. No. 3.9e-65;
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;

QY 1 ATGGCTCGTACGAAGCAAACTGCAAGAAATCACACGAGGAAAGCTCCGAGGACTCTG 60
Db 62 ATGGCTCGTACGAAGCAGACTGTAGGAAATCCACCGGAGGAAAGCTCCGAGGACAA 121
QY 61 CTGCTACCAAGCGGCGAGGAAATCTGCGCGACTACTGAGGAGTCAAGAACTCTCAC 120
Db 122 CTGCTACCAAGCGGCGAGGAAATCTGCGCGACTACTGAGGAGTCAAGAACTCTCAC 181
QY 121 CGTTACCGTCCCGAAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGACACAGAG 180
Db 182 CGTTACCGTCCCGAAACCGTCTCTTCGTGAGATTCGTAATACCAAGAGACACAGAG 241
QY 181 TTGTTGATCCGTAAACTCTCTTTTCAACGCTCTGTTGTAATCGCTCAAGATTACAAG 240
Db 242 TTGTTGATCCGTAAAGCTCTCGTTTTCAGCGTCTGTTGTAATCGCTCAAGATTACAAG 300
QY 241 ACGGATCTGAGATTCAGAGCCATCGGTGTTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
Db 301 ----- 300
QY 301 TTGACGATCTGAGATTCAGAGCCATCGGTGTTAGCTCTTCAAGAGCTGCGAGGCA 360
Db 301 --GACAGACCTGAGTTCAGAGCCACGACGTTGTCGCTTTCAGGAGCTGCGAGGCT 358
QY 361 TATTGTTGGTGTGTTGTTGAAGACACAAATCTGTGTCCTTCAATGCAAGAGGTTACG 420
Db 359 TATTGTTGGTGTGTTGTTGAAGACACAAATCTGTGTCCTTCAATGCAAGAGGTTACG 418
QY 421 ATAATGCTTAAAGATGTTCAATTCGCAAGAGGATTCGTGAGAGCGTCTTAG 474
Db 419 ATCATGCTTAAAGATGTTCAATTCGCTAGGCTATTCTGTTGAGAGCGTCTTAG 472

```

```

RESULT 9
CD823357      676 bp      mRNA      linear      EST 10-JUL-2003
LOCUS         BN25.048106F020110 BN25 Brassica napus cDNA clone BN25048106, mRNA
DEFINITION   sequence.
ACCESSION    CD823357
VERSION      CD823357.1 GI:32505297
KEYWORDS     EST.
SOURCE       Brassica napus (rape)
ORGANISM     Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE    1 (bases 1 to 676)
AUTHORS      Genoplante.
TITLE        Genoplante, a major partnership french program in plant genomics
JOURNAL      Unpublished (2003)
COMMENT      Contact: Genoplante
              Genoplante
              93, rue Henri Rochefort 91025 EVRY CEDEX France
              Tel: 33 1 69 47 54 00
              Fax: 33 1 69 47 54 10
              This sequence has been generated in the framework of the french
              plant genomics programme 'Genoplante' (http://www.genoplante.com
              and http://genoplante-info.inbio.gen.fr).
              Location/Qualifiers
              source          1. .676

```


/organism="Brassica napus"
/mol_type="mRNA"
/cultivar="Jet neuf"
/db_xref="taxon:3708"
/clone="BN25048106"
/tissue_type="seed"
/clone_lib="BN25"

ORIGIN

Query Match 54.1%; Score 256.4; DB 14; Length 676;
Best Local Similarity 75.9%; Pred. No. 4e-65;
Matches 360; Conservative 0; Mismatches 51; Indels 63; Gaps 1;
QY 1 ATGGCTCGTACGAGCAAACTCGAAGAAATCACGAGGAGAAAGCTCCGAGGACTCTG 60
DB 63 ATGGCTCGTACGAGCAGACTGCTAGAAATCACCGGAGGAGAAAGCTCCGAGGAGCAA 122
QY 61 CTCGCTTACCAAGCGCGGAGGAAATCTGCGCGACTACTGGAGGAGTCAAGAAACCTCAC 120
DB 123 CTCGCCACCAAGCGCGGAGGAAATCGGCGCCGACACCGGAGGAGTCAAGAGCCTCAC 182
QY 121 CATTACCGTCCCGGAGCCGCTCTCTTCGTGAGATTCTGTAATACCAAGAGACACAGAG 180
DB 183 CGTTACCGTCCCGGAGCCGCTCTCTTCGTGAGATTCTGTAATACCAAGAGACACAGAG 242
QY 181 TTGTTGATCGGTAACCTTCTTTCAACGCTTGTTCGTGAATCGCTCAAGATTACAG 240
DB 243 TTGTTGATCGGTAACCTTCTTTCAACGCTTGTTCGTGAATCGCTCAAGATTACAG 301
QY 241 ACGGATCTGAGATTCCAGAGCCATCGGTGTTAGCTCTTCAAGAGCTGCTGAAGCATAT 300
DB 302 ----- 301
QY 301 TTGACGGAATCTGAGATTCCAGAGCCATCGGTGTTAGCTCTTCAAGAGCTCGGAGGCA 360
DB 302 --GACAGACTGAGGTTCCAGAGCCAGCAGTGTTCGCGCTTCAAGAGCTCGGAGGCT 359
QY 361 TATTTGGTGGTTTGTGTAAGACACAAATCTCTGTGCCATTCATCAAGAGAGGTTACG 420
DB 360 TATCTGGTGGTTTGTGTAAGACACAAATCTCTGTGCCATTCATCAAGAGAGGTTACG 419
QY 421 ATAAATGCTAAAGATGTTCAATTTGGCAAGAGGATTCTGTGAGAGCGGTGTTAG 474
DB 420 ATCAATGCTAAAGATGTTCAATTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 473

RESULT 10
AA650659 527 bp mRNA linear EST 31-OCT-1997
LOCUS 30773 Lambda-PRL2 Arabidopsis thaliana cdna clone 276G8T7, mRNA
DEFINITION
AA650659
ACCESSION
VERSION AA650659.1 GI:2580751
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 527)
Newman, T., deBruijn, F.J., Green, P., Keegstra, K., Kende, H.,
McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomas, M.,
Retzel, E. and Somerville, C.
Genes Galore: a summary of methods for accessing results from
large-scale partial sequencing of anonymous Arabidopsis cdna clones
Plant Physiol. 106, 1241-1255 (1994)
95148729
MEDLINE
PUBMED 7846151
CONTACT: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI

Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@bm.cl.msu.edu
Seq primer: T7 dye primer.

FEATURES

Location/Qualifiers
1..527
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="var columbia"
/db_xref="taxon:3702"
/clone="276G8T7"
/clone_lib="Lambda-PRL2"
/note="Vector: lambda Zip-Lox; Site 1: Sal; Site 2: Not;
Lambda PRL2 is a cdna library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark- rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques. The vector is BRL's lambda Zip-Lox. The cdna
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cdna."

ORIGIN

Query Match 53.9%; Score 255.4; DB 9; Length 527;
Best Local Similarity 80.2%; Pred. No. 7.3e-65;
Matches 341; Conservative 0; Mismatches 21; Indels 63; Gaps 1;
QY 50 CGAGGACTCTGCTCTACCAAGCGGAGGAAATCTGCGCGACTACTGAGAGAGTCA 109
DB 1 CGAGCAAGCAGCTCGCTACCAAGCGGCAAGGAATCTGCACCGACTACCGGAGGAGTCA 60
QY 110 AGAAACCTCACCGTTACCGTCCCGGACCGTCTTCTGAGATTCTGTAATACCAGA 169
DB 61 AGAAACCTCACCGTTACCGTCTGGAACCGTCTTCTGAGATTCTGTAATACCAGA 120
QY 170 AGAGCACAGAGTGTGTATCCCGTAAACTCTTTCAAGCTTGTTCGTGAAATCGCTC 229
DB 121 AGAGCACAGAGTGTGTATCCCGTAACTTCATTCACAGCTTGTTCGTGAAATCGCTC 180
QY 230 AGATTACAAGACCGGATCTGAGATCCAGAGCCATCGGTGTTAGCTCTTCAAGAGCTG 289
DB 181 AAGATTTCAA----- 190
QY 290 CTGAACCATATTGTGCGGATCTGAGATCCAGAGCCATCGGTGTTAGCTCTTCAAGAG 349
DB 191 -----GACGATCTGAGATCCAAAGCCATGAGTGTAGCTCTTCAAGAG 237
QY 350 CTGCGGAGCATATTGTGGGTTGTTTGAAGACACAAATCTGTGTGCATTTCATGCAA 409
DB 238 CTGCGGAGCATATTGTGGTGGTTGTTTGAAGACACAAATCTGTGTGCATTTCATGCAA 297
QY 410 AGAGGTTACGATATGCTTCAAGATGTTCAATGCGCAAGAGGATTCGTGGAGAGCTG 469
DB 298 AGAGGTTACGATATGCTTCAAGATGTTCAATGCGCAAGAGGATTCGTGGAGAGCTG 357
QY 470 CTTAG 474
DB 358 CTTAG 362

RESULT 11
BQ791478 545 bp mRNA linear EST 30-JUL-2002
LOCUS BQ791478
DEFINITION E4401 Chinese cabbage etiolated seedling library Brassica rapa
subsp. pekinensis cdna clone E4401, mRNA sequence.
ACCESSION BQ791478
VERSION BQ791478.1 GI:22006440
KEYWORDS EST.
SOURCE Brassica rapa subsp. pekinensis
ORGANISM Brassica rapa subsp. pekinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 545)
Ryu,S.H., Yang,K.A., Lee,S.Y., Kim,H.-I., Cho,M.J. and Lim,C.O.
Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
(2002)
Unpublished (2002)
JOURNAL
COMMENT
Contact: Lim, C.O.
Plant Molecular Biology & Biotechnology Research Centre
Gyeongsang National University
#900 Gazwa-dong, Jinju 660-701, Korea
Tel: 82 55 751 6255
Fax: 82 55 759 9363
Email: colim@nongae.gsnu.ac.kr
Seq primer: T7
Location/Qualifiers
1..545
/organism="Brassica rapa subsp. pekinensis"
/mol_type="mRNA"
/cultivar="Jangwon"
/sub_species="pekinensis"
/db_xref="taxon:51351"
/clone="E4401"
/tissue_type="Etiolated seedling"
/lab_host="XL-1 Blue"
/clone_lib="Chinese cabbage etiolated seedling library"
/note="Vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"

ORIGIN
Query Match 50.2%; Score 237.8; DB 13; Length 545;
Best Local Similarity 73.6%; Pred. No. 1.3e-59;
Matches 348; Conservative 0; Mismatches 62; Indels 63; Gaps 1;
QY 1 ATGGCTCTGACCAAGCAAACTCAGAAATCACACGGAGGAAGAGCTCCGAGGACTCTG 60
Db |||||||
QY 34 ATGGCTCTGACCAAGCAAACTCAGAAATCACACGGAGGAAGAGCTCCGAGGAGCAG 93
Db |||||||
QY 61 CTCGCTACCAAGCGCGGAGGAAATCTGCGCGACTACTGAGGAGTCAAGAAACCTCAC 120
Db |||||||
QY 94 CTCGCCCAAGCGCGGAGGAAATCTGCTCCGACCACCGGAGGAGTCAAGAAACCTCAC 153
QY 121 CGTTACCGTCCCGGAACCGTCTCTCTGAGATTCGTAATACAGAGAGCAGAG 180
Db |||||||
QY 154 CGTTACCGTCCCGGAACCGTCTCTCTGAGATTCGTAATACAGAGAGCAGAG 213
QY 181 TTGTTGATCCGTAACCTCTCTTCAACCTCTGTTGTAATCGCTCAAGATTACAG 240
Db |||||||
QY 214 CTTTGAUCCGTAAGCTTCCCTTCAGCGTCTGTTGTAATCGCTCAGGACTTCAG 273
QY 241 ACGGATCTGAGATTCAGAGCCATCGGTTGTTAGCTTTCAAGAAAGCTCTGAAGCATAT 300
Db |||||||
QY 274 ACGGATCTGAGATTCAGAGCCATCGGTTGTTAGCTTTCAAGAAAGCTCTGAAGCATAT 324
QY 301 TTGACGGATCTGAGATTCAGAGCCATCGGTTGTTAGCTTTCAAGAAAGCTCGGAGGCA 360
Db |||||||
QY 325 -----GAGGCG 330
QY 361 TATTTGGTGGGTTGTTTGAAGACAAATCTGTGTGCCATTATGCAAGAGGGTTACG 420
Db |||||||
QY 331 TACCTTGGGTTGTTTGAAGATCTATCTCTGCGGATTCACGCGAGAGGGTGACG 390
QY 421 ATAATGCCCTAAGATTTCAATTTGGCAAGAGGATTCGTGGAGCGTCTTGA 473
Db |||||||
QY 391 ATCATGCCCAAGGATATTAGTTGGGAGGAGGATTCGTGGTGAACGTCTTGA 443

RESULT 12
AV552968
LOCUS
DEFINITION
cDNA clone RZ50d08R 5', mRNA sequence.
ACCESSION
AV552968
VERSION
AV552968.1 GI:8724381
KEYWORDS
EST.

rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 545)
Ryu,S.H., Yang,K.A., Lee,S.Y., Kim,H.-I., Cho,M.J. and Lim,C.O.
Expressed Sequence Tags of Chinese Cabbage Etiolated Seedling cDNA
(2002)
Unpublished (2002)
JOURNAL
COMMENT
Contact: Lim, C.O.
Plant Molecular Biology & Biotechnology Research Centre
Gyeongsang National University
#900 Gazwa-dong, Jinju 660-701, Korea
Tel: 82 55 751 6255
Fax: 82 55 759 9363
Email: colim@nongae.gsnu.ac.kr
Seq primer: T7
Location/Qualifiers
1..545
/organism="Brassica rapa subsp. pekinensis"
/mol_type="mRNA"
/cultivar="Jangwon"
/sub_species="pekinensis"
/db_xref="taxon:51351"
/clone="E4401"
/tissue_type="Etiolated seedling"
/lab_host="XL-1 Blue"
/clone_lib="Chinese cabbage etiolated seedling library"
/note="Vector: pSPORT 1; Site_1: Sal I; Site_2: Not I"

ORIGIN
Query Match 50.2%; Score 237.8; DB 13; Length 545;
Best Local Similarity 73.6%; Pred. No. 1.3e-59;
Matches 348; Conservative 0; Mismatches 62; Indels 63; Gaps 1;
QY 1 ATGGCTCTGACCAAGCAAACTCAGAAATCACACGGAGGAAGAGCTCCGAGGACTCTG 60
Db |||||||
QY 34 ATGGCTCTGACCAAGCAAACTCAGAAATCACACGGAGGAAGAGCTCCGAGGAGCAG 93
Db |||||||
QY 61 CTCGCTACCAAGCGCGGAGGAAATCTGCGCGACTACTGAGGAGTCAAGAAACCTCAC 120
Db |||||||
QY 94 CTCGCCCAAGCGCGGAGGAAATCTGCTCCGACCACCGGAGGAGTCAAGAAACCTCAC 153
QY 121 CGTTACCGTCCCGGAACCGTCTCTCTGAGATTCGTAATACAGAGAGCAGAG 180
Db |||||||
QY 154 CGTTACCGTCCCGGAACCGTCTCTCTGAGATTCGTAATACAGAGAGCAGAG 213
QY 181 TTGTTGATCCGTAACCTCTCTTCAACCTCTGTTGTAATCGCTCAAGATTACAG 240
Db |||||||
QY 214 CTTTGAUCCGTAAGCTTCCCTTCAGCGTCTGTTGTAATCGCTCAGGACTTCAG 273
QY 241 ACGGATCTGAGATTCAGAGCCATCGGTTGTTAGCTTTCAAGAAAGCTCTGAAGCATAT 300
Db |||||||
QY 274 ACGGATCTGAGATTCAGAGCCATCGGTTGTTAGCTTTCAAGAAAGCTCTGAAGCATAT 324
QY 301 TTGACGGATCTGAGATTCAGAGCCATCGGTTGTTAGCTTTCAAGAAAGCTCGGAGGCA 360
Db |||||||
QY 325 -----GAGGCG 330
QY 361 TATTTGGTGGGTTGTTTGAAGACAAATCTGTGTGCCATTATGCAAGAGGGTTACG 420
Db |||||||
QY 331 TACCTTGGGTTGTTTGAAGATCTATCTCTGCGGATTCACGCGAGAGGGTGACG 390
QY 421 ATAATGCCCTAAGATTTCAATTTGGCAAGAGGATTCGTGGAGCGTCTTGA 473
Db |||||||
QY 391 ATCATGCCCAAGGATATTAGTTGGGAGGAGGATTCGTGGTGAACGTCTTGA 443

RESULT 12
AV552968
LOCUS
DEFINITION
cDNA clone RZ50d08R 5', mRNA sequence.
ACCESSION
AV552968
VERSION
AV552968.1 GI:8724381
KEYWORDS
EST.

SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 545)
Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation
of 12,028 non-redundant expressed sequence tags from normalized and
size-selected cDNA libraries
DNA Res. 7 (3), 175-180 (2000)
20363093
MEDLINE
PUBMED
10907847
COMMENT
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yazusa 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1..545
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZ50d08R"
/tissue_type="roots"
/clone_lib="Arabidopsis thaliana roots Columbia"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
XhoI"

ORIGIN
Query Match 49.2%; Score 233; DB 9; Length 545;
Best Local Similarity 72.9%; Pred. No. 3.4e-58;
Matches 345; Conservative 0; Mismatches 63; Indels 63; Gaps 1;
QY 1 ATGGCTCTGACCAAGCAAACTCAGAAATCACACGGAGGAAGAGCTCCGAGGACTCTG 60
Db |||||||
QY 26 ATGGCTCTGACCAAGCAAACTCAGAAATCACACGGAGGTAAGTCCCGAGGAGCAG 85
QY 61 CTCGCTACCAAGCGCGGAGGAAATCTGCGCGACTACTGAGGAGTCAAGAAACCTCAC 120
Db |||||||
QY 86 CTTGCCCAAGCGCGGAGGAAATCAGCAACACCGGAGGAGTCAAGAAACCTCAC 145
QY 121 CTTACCGTCCCGGAACCGTCTCTCTGAGATTCGTAATACCAAGAGCACAG 180
Db |||||||
QY 146 CGTTACCGTCCAGAACCGTCTCTCTGAGATTCGTAATACCAAGAGTACTGA 205
QY 181 TTGTTGATCCGTAACCTCTCTTTCAAGCTCTTGTGTAATCGCTCAAGATTACAG 240
Db |||||||
QY 206 TTGTTGATCCGCAAGCTTCCCTTCCAGCGTCTCTGTTCCGGAATAGCTCAGGACTTCAAG 265
QY 241 ACGGATCTGAGATTCAGAGCCATCGGTTGTTAGCTTTCAAGAGCTCTGAAGCATAT 300
Db |||||||
QY 266 ACGGATCTGAGATTCAGAGTCAAGCGGTTGTTAGCTTTCAAGAGGCGCT----- 316
QY 301 TTGACGGATCTGAGATTCAGAGCCATCGGTTGTTAGCTTTCAAGAAAGCTCGGAGGCA 360
Db |||||||
QY 317 -----GAGGCA 322
QY 361 TATTTGGTGGGTTGTTTGAAGACAAATCTGTGTGCCATTATGCAAGAGGGTTACG 420
Db |||||||
QY 323 TACCTAGTTGGACTTTTCGAGGATACTAATCTCTCGCTATTTCAGCCCAAGAGGGTTACG 382
QY 421 ATAATGCCCTAAGATTTCAATTTGGCAAGAGGATTCGTGGAGCGTCTTGA 473
Db |||||||
QY 383 ATCATGCCCTAAGGACATTCATTTGGCTAGAGATTCGTGGAGAACGTCCTTA 435

RESULT 13
CD830234
LOCUS
DEFINITION
cDNA clone BN40.044M18F011221 BN40 Brassica napus cDNA clone BN40044M18, mRNA
sequence.
CD830234
BN40.044M18F011221 BN40 Brassica napus cDNA clone BN40044M18, mRNA
sequence.
EST 10-JUL-2003

```

ACCESSION   CD830234
VERSION     CD830234.1  GI:32512174
KEYWORDS    EST.
SOURCE      Brassica napus (rape)
ORGANISM    Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 603)
AUTHORS     Genoplante.
            Genoplante, a major partnership french program in plant genomics
            Unpublished (2003)
TITLE       Genoplante
JOURNAL
COMMENT     Contact: Genoplante
            Location/Qualifiers
FEATURES             source
             1..603
                /organism="Brassica napus"
                /mol_type="mRNA"
                /cultivar="Jet neuf"
                /db_xref="taxon:3708"
                /clone="BN4004M18"
                /tissue_type="seed"
                /clone_lib="BN40"

ORIGIN
Query Match      49.2%; Score 233; DB 14; Length 603;
Best Local Similarity 72.9%; Pred. No. 3.6e-58;
Matches 345; Conservative 0; Mismatches 65; Indels 63; Gaps 1;

QY      1  ATGGCTCGTAGAGCAAACTGCAAGAAATACACAGGAGGAAAGCTCCGAGGACTCTG 60
DB      69  ATGGCTCGTAGAGCAAAACCGCTAGAAAGTCCACGAGGAGCAAGTCCGAGGAGCAG 128
QY      61  CTCGCTACCAAGGGCGGAGAAATCTGCGCGGACTACTGAGGAGGTCAAGAAACCTTAC 120
DB      129  CTCGCCACCAAGGGCGGAGAAATCTGCTCCGACGACCGGAGGAGTCAAGAGCCGCAC 188
QY      121  CGTTACCGTCCGAGAACCGTCTGCTGAGATTGTTAAATACCAAGAGGACACAGAG 180
DB      189  CGTTACCGTCCGAGAACCGTCTGCTGAGATCAGAGATCAGAGAGGACCGAG 248
QY      181  TTGTTGATCCGTAACCTTCCTTTTCAAGCTCTTGTGTAATCGCTCAAGATTACAAG 240
DB      249  CTCCTTGATCCGTAGCTTCTTTCCAGCGTCTGTTGTTGTAATCGCTCAGGACTTCAA- 307
QY      241  ACGGATCTGAGATTCAGAGCCATGCGGTGTAGTCTCTTCAAGAGCTGCTGAAGCATAT 300
DB      308  ----- 307
QY      301  TTGACGGATCTGAGATTCAGAGCCATGCGGTGTAGTCTCTTCAAGAGCTGCGAGGCA 360
DB      308  --GACGATCTGAGATTCAGAGCCATGCGGTGTAGTCTCTTCAAGAGGCGCGGAGGCG 365
QY      361  TATTGTGGGTGTTGTTTGAAGACACAAATCTGTGTGCATTCATGCAAGAGGTTACG 420
DB      366  TACCTCGTGGGTTGTTTCGAGGATCTAATCTCTGCGGATTCACGCGAAGAGGTCACG 425
QY      421  ATATGCTTAAGATGTTCAATTGCAAGAGGATTCGTGAGAGCGTGCTTA 473
DB      426  ATCATGCCCAAGGATATTTCAGTTGGCCAGGAGGATTCGTGTAACGCTGCTTA 478

RESULT 14
AI998977/c 611 bp mRNA linear EST 08-SEP-1999
LOCUS      701547802 A. thaliana, Columbia Col-0, rosette-2 Arabidopsis
DEFINITION thaliana cDNA clone 701547802, mRNA sequence.

```

```

ACCESSION   AI998977
VERSION     AI998977.1  GI:5845882
KEYWORDS    EST.
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1 (bases 1 to 611)
AUTHORS     Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,
            Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P.,
            Gorsone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D.,
            Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C.,
            Carpio, T., Policky, J., Suzuki, G., Argentin, C., Shah, S., and
            Nobrega, A., Murry, L., Turner, C., Krikorian, S., Elder, L., and
            Hanson, D.
            Arabidopsis thaliana Gene Expression MicroArray
            Unpublished (1999)
TITLE       Arabidopsis thaliana
JOURNAL
COMMENT     Contact: David Smoller, Ph.D.
            Genome Systems, Inc., a wholly owned subsidiary of Inocyte
            Pharmaceuticals, Inc.
            4633 World Parkway Circle, St. Louis, MO 63134, USA
            Tel: 877-577-2733
            Fax: 314-427-3324
            Email: service@genomesystems.com.
            Location/Qualifiers
FEATURES             source
             1..611
                /organism="Arabidopsis thaliana"
                /mol_type="mRNA"
                /cultivar="Columbia Col-0"
                /db_xref="taxon:3702"
                /clone="701547802"
                /tissue_type="rosette"
                /dev_stage="4 - 7 weeks"
                /clone_lib="A. thaliana, Columbia Col-0, rosette-2"
                /notes="Vector: pSPORT; Site 1: NotI; Site 2: SalI; cDNA
                library was derived from untreated rosette tissue from
                Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks.
                Plants were grown in 1:1 peat moss/vermiculite/perlite
                soil at 22 deg. C +/- 3 deg. C under constant light, and
                watered with fertilizer. cDNA synthesis was initiated
                using a NotI-oligo(dT) primer. Double-stranded cDNA was
                blunted, ligated to SalI adaptors, digested with NotI,
                size-selected, and cloned into the NotI and SalI sites of
                the pSPORT vector."

ORIGIN
Query Match      49.2%; Score 233; DB 9; Length 611;
Best Local Similarity 72.9%; Pred. No. 3.6e-59;
Matches 345; Conservative 0; Mismatches 65; Indels 63; Gaps 1;

QY      1  ATGGCTCGTAGAGCAAACTGCAAGAAATACACAGGAGGAAAGCTCCGAGGACTCTG 60
DB      547  ATGGCTCGTAGAGCAAAACCGCTAGAAATCACCAGGAGGTAAAGTCCCGAAGCAG 488
QY      61  CTCGCTACCAAGGGCGGAGAAATCTGCGCGGACTACTGCGGAGGTCAAGAAACCTTAC 120
DB      487  CTTGCCACCAAGGGCGGAGAAATCAGACCAACACCGGAGGAGTCAAGAGCCTTAC 428
QY      121  CGTTACCGTCCGAGAACCGTCTGCTGAGATTGTTAAATACCAAGAGGACACAGAG 180
DB      427  CGTTACCGTCCGAGAACCGTCTGCTGAGATTGTTAAATACCAAGAGGAGTACTGAA 368
QY      181  TTGTTGATCCGTAAACCTTCCTTTTCAAGCTCTTGTGTAATCGCTCAAGATTACAAG 240
DB      367  TTGTTGATCCGTAAACCTTCCTTTTCAAGCTCTTGTGTAATCGCTCAAGATTACAAG 308
QY      241  ACGGATCTGAGATTCAGAGCCATGCGGTGTAGTCTCTTCAAGAGCTGCTGAAGCATAT 300
DB      307  ACGGATCTGAGGTTTCAGAGTCACGCGGTGTAGTCTCTTCAAGAGGCGGCT----- 257
QY      301  TTGACGATCTGAGATTCCAGAGCCATGCGGTGTAGTCTCTTCAAGAGCTGCGGAGGCA 360

```

```

Db      256 -----GAGGCA 251
Qy      361 TATTGGTGGTTTGTGTAAGACACAAATCTGTGTCATTCATGCAAGAGGTTACG 420
Db      250 TACCTAGTTGACTTTTCGAGGACTAAATCTCGCTATTACGCCAAGAGGTTACG 191
Qy      421 ATAATGCTTAAGATGTTCAATTGGCAGAGAGGATTCTGTGAGAGCGTGCTTA 473
Db      190 ATCATGCCCTAAGGACATTCAAATTGGCTAGAGAAATTCGTGGAGACGTGCTTA 138

RESULT 15
CD822585
LOCUS   BN25.045M06F020108 BN25 Brassica napus cDNA clone BN25045M06, mRNA
DEFINITION
sequence.
ACCESSION   CD822585
VERSION     CD822585.1 GI:32504525
KEYWORDS    EST.
SOURCE      Brassica napus (rape)
ORGANISM    Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 643)
AUTHORS     Genoplante.
TITLE       Genoplante, a major partnership french program in plant genomics
JOURNAL     Unpublished (2003)
COMMENT     Contact: Genoplante
Genoplante
93, rue Henri Rochefort 91025 EVRY CEDEX France
Tel: 33 1 69 47 54 00
Fax: 33 1 69 47 54 10
This sequence has been generated in the framework of the french
plant genomics programme 'Genoplante' (http://www.genoplante.com
and http://genoplante-info.infobiogen.fr).

FEATURES             source
1..643
    /organism="Brassica napus"
    /mol_type="mRNA"
    /cultivar="Jet neuf"
    /db_xref="taxon:3708"
    /clone="BN25045M06"
    /tissue_type="seed"
    /clone_lib="BN25"

ORIGIN
Query Match      49.2%; Score 233; DB 14; Length 643;
Best Local Similarity 72.9%; Pred. No. 3.7e-58;
Matches 345; Conservative 0; Mismatches 65; Indels 63; Gaps 1;

Qy      1 ATGGTCTGTACGACGAACCTCAGAAATCACAACGGAGGAAAGCTCCGAGGACTCTG 60
Db      69 ATGGCTGTGTACGAAACAAACCGCTAGAAAGTCCACGGAGGCAAAAGCTCCCGAAGCAG 128
Qy      61 CTCGTCTACCAAGCGCGAGGAAATCTGCGCGACTACTGGAGGAGTCAAGAAACCTCAC 120
Db      129 CTCGCCACCAAGCGCGAGGAAATCTGTCGACGACCGGAGGAGTCAAGAGCGGCAC 188
Qy      121 CATTACCGTCCCGGAACCGTCTCTTCTGTGAGATTCGTAATACAGAAAGACAGAG 180
Db      189 CGTTACCGTCCAGGAACCGTCTGCGCTTCGTGAGATCAGGAAGTACCAGAAAGACGCGAG 248
Qy      181 TTGTTGATCCGTAAATCTCTTTCAACGCTTGTCTGTAATCGCTCAAGATTACAG 240
Db      249 CTCTTGATCCGTAAGCTCTCTTTCCAGCGTCTGTTCTGTAATCGCTCAGGACTTCAA- 307
Qy      241 ACGGATCTGAGATTCACAGGCCATGCGGTGTAGCTCTTCAAGAGAGCTGCTGAAGCATAT 300
Db      308 -----
Qy      301 TTGACGGATCTGAGATTCAGAGCCATGCGGTGTAGCTCTTCAAGAGAGCTGCGAGGCA 360

```

```

Db      308 --GACGGATCTGAGTTCCAGAGCCACGCCGTGTTGGCTCTCCAGGAGGCGCGAGGCG 365
Qy      361 TATTGGTGGTTTGTGTAAGACACAAATCTGTGTCATTCATGCAAGAGGTTACG 420
Db      366 TACCTCGTGGTTTGTTCGAGGACTAAATCTCTGCGGATTCACGCCAAGAGGTTACG 425
Qy      421 ATAATGCTTAAGATGTTCAATTGGCAGAGAGGATTCTGTGAGAGCGTGCTTA 473
Db      426 ATCATGCCCAAGGATATTCAGTTGGCCAGGAGGATTCGTGTAACGTGCTTA 478

```

Search completed: June 28, 2004, 18:01:36
Job time : 1648.18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 18:04:41 ; Search time 43 seconds
(without alignments)
364.633 Million cell updates/sec

Title: AAC50492

Perfect score: 852

Sequence: 1 MADBEKLPFGWEKMRSSG.....GEMSGPVFTDGGIHLIRTE 163

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	852	100.0	163	2	S68520	Pin1 protein - hum
2	814	95.5	165	2	JC7136	peptidylprolyl iso
3	456.5	53.6	166	2	T08426	Pin1 protein, homol
4	388.5	45.6	175	2	T41093	peptidyl-prolyl ci
5	356.5	41.8	170	2	S2764	ESSI protein - yea
6	301	35.3	119	2	E84559	probable peptidyl-
7	180.5	21.2	292	2	S15269	post-translocation
8	177	20.8	247	2	A96934	peptidyl-prolyl ci
9	158	18.5	142	2	E86392	hypothetical prote
10	157.5	18.5	126	2	T31601	hypothetical prote
11	153.5	18.0	331	2	A3267	peptidyl-prolyl ci
12	149	17.5	273	2	S2412	major antigenic pe
13	148.5	17.4	333	2	A83797	protein secretion
14	147	17.3	655	2	E82712	peptidyl-prolyl ci
15	145	17.0	347	2	A81794	probable rotamase
16	145	17.0	348	2	B81216	peptidyl-prolyl ci
17	138.5	16.3	621	2	A83421	peptidyl-prolyl ci
18	138	16.2	93	2	E91217	peptidylprolyl iso
19	138	16.2	93	2	F85063	peptidylprolyl iso
20	138	16.2	93	2	S48658	peptidylprolyl iso
21	137	16.1	93	2	A80923	peptidyl-prolyl ci
22	127.5	15.0	320	2	H89970	hypothetical prote
23	127	14.9	93	2	A80471	peptidylprolyl iso
24	126.5	14.8	92	2	F83395	peptidyl-prolyl ci
25	124	14.6	300	2	A92989	peptidyl-prolyl ci
26	124	14.6	315	2	C98294	cell binding facto
27	120.5	14.1	92	2	A82396	peptidyl-prolyl ci
28	119.5	14.0	434	2	A00061	peptidylprolyl iso
29	117	13.7	293	2	AC1352	post-translocation

30	116	13.6	428	2	B90636	survival protein [
31	116	13.6	428	2	B85487	survival protein [
32	116	13.6	428	2	B84726	probable peptidylp
33	114.5	13.4	431	2	D82323	survival protein S
34	114.5	13.4	708	2	I83196	NEDD-4 ORF - mouse
35	114	13.4	93	2	C83123	peptidyl-prolyl ci
36	114	13.4	291	2	AF1722	post-translocation
37	112.5	13.2	266	2	S02510	nifM protein - Kle
38	112	13.1	820	2	T46412	ubiquitin-protein
39	110.5	13.0	293	2	H43706	nifM protein - Azo
40	109.5	12.9	221	2	T52622	probable peptidylp
41	109	12.8	428	2	AH0513	survival protein S
42	108.5	12.7	353	2	D97295	parvulin-like pept
43	108	12.7	336	2	D70113	basic membrane pro
44	107.5	12.6	120	2	T26779	hypothetical prote
45	107.5	12.6	887	2	S70642	ubiquitin ligase N

ALIGNMENTS

RESULT 1

S68520

Pin1 protein - human

C:Species: Homo sapiens (man)

C:Date: 15-Nov-1996 #sequence_revision 27-Feb-1997 #text_change 29-Sep-1999

C:Accession: S68520

R:Idu, K.F.; Hanes, S.D.; Hunter, T.

Nature 380, 544-547, 1996

A:Title: A human peptidyl-prolyl isomerase essential for regulation of mitosis.

A:Reference number: S68520; MUID:96195064; PMID:8606777

A:Accession: S68520

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-163 <LUK>

A:Cross-references: EMBL:U49070; NID:gl332709; PIDN:AAC50492.1; PID:gl332710

C:Genetics:

A:Gene: GDB:PIN1; dod

A:Cross-references: GDB:5218381

C:Superfamily: Yeast ESSI protein; WW repeat homology

F:5-43/Domain: WW repeat homology <WW1>

Query Match	100.0%	Score 852;	DB 2;	Length 163;
Best Local Similarity	100.0%	Pred. No. 9.4e-72;		
Matches 163;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	MADBEKLPFGWEKMRSSGRVYFNHITNASQWERPSGSSGGKNGQGEPARVRCSHL	60
Db	1	MADBEKLPFGWEKMRSSGRVYFNHITNASQWERPSGSSGGKNGQGEPARVRCSHL	60

QY	61	LVKHSQRRPSSWRQEKITRTKEALELINGYIOKIKSGDEDFESLASOFSDCSSAKARG	120
Db	61	LVKHSQRRPSSWRQEKITRTKEALELINGYIOKIKSGDEDFESLASOFSDCSSAKARG	120

QY	121	DLGAFSRGQMKPPEDASFALRTGEMSGPVFTDGGIHLIRTE	163
Db	121	DLGAFSRGQMKPPEDASFALRTGEMSGPVFTDGGIHLIRTE	163

RESULT 2

JC7136

peptidylprolyl isomerase (EC 5.2.1.8) - mouse

N:Alternate names: parvulins, pin1 protein

C:Species: Mus musculus (house mouse)

C:Date: 04-Mar-2000 #sequence_revision 04-Mar-2000 #text_change 18-Aug-2000

C:Accession: JC7136

R:Fujimori, F.; Takahashi, K.; Uchida, C.; Uchida, T.

Biochem. Biophys. Res. Commun. 265, 658-663, 1999

A:Title: Mice lacking Pin1 develop normally, but are defective in entering cell cycle f:

A:Reference number: JC7136; MUID:20070807; PMID:10600477

A:Accession: JC7136

A:Molecule type: mRNA

A:Residues: 1-165 <FUJ>

A;Cross-references: DBJ:AB009691; NID:g6468199; PIDN:BAA7037.1; PID:g6468200
C;Comment: This protein is a member of peptidyl prolyl cis/trans isomerase family. It is
tes in oligopeptides, controls G2/M transition of HeLa cells, regulates entry into G1 an
C;Genetics:
A;Gene: Pin1
C;Superfamily: yeast ESS1 protein; WW repeat homology
C;Keywords: cis-trans-isomerase; mitosis
F;5-43/Domain: WW repeat homology <WWR>

Query Match 95.5%; Score 814; DB 2; Length 165;
Best Local Similarity 95.2%; Pred. No. 3.2e-68;
Matches 157; Conservative 3; Mismatches 3; Indels 2; Gaps 1;
QY 1 MADEKLPQGWKMRSSGRVYFNHITNASQWRPSSGG--KNGQGEPAVRCS 58
DB 1 MADEKLPQGWKMRSSGRVYFNHITNASQWRPSSGGTGGSSKNGQGEPAVRCS 60
QY 59 HLLVKSQRRRPSWQEKITRKEALELINGYIQIKSGEDFSLASQFSDCSSAK 118
DB 61 HLLVKSQRRRPSWQEKITRKEALELINGYIQIKSGEDFSLASQFSDCSSAK 120
QY 119 RGDLPFSGQMQKPPEDAFALRTGEMSGPVFTDGGIHLIRTE 163
DB 121 RGDLPFSGQMQKPPEDAFALRTGEMSGPVFTDGGIHLIRTE 165

RESULT 3
T08426
Pin1 protein homolog dodo - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 17-Nov-2000
C;Accession: T08426
R;Maleszka, R.; de Couet, H.G.; Miklos, G.L.
Proc. Natl. Acad. Sci. U.S.A. 95, 3731-3736, 1998
A;Title: Data transferability from model organisms to human beings: insights from the fu
A;Reference number: Z16415; MUID:98188272; PMID:9520435
A;Accession: T08426
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-166 <MAL>
A;Cross-references: EMBL:AF01777; NID:g3004652; PIDN:AAC28408.1; PID:g3004661
A;Experimental source: strain Cantons
C;Genetics:
A;Gene: dod
A;Cross-references: FlyBase:FBgn0024251
A;Introns: 20/1; 131/1
C;Superfamily: yeast ESS1 protein; WW repeat homology
F;5-43/Domain: WW repeat homology <WWR>

Query Match 53.6%; Score 456.5; DB 2; Length 166;
Best Local Similarity 56.7%; Pred. No. 4.5e-35;
Matches 93; Conservative 16; Mismatches 52; Indels 3; Gaps 2;

QY 1 MADEKLPQGWKMRSSGRVYFNHITNASQWRPSS--GNSSSGKNGQGE-PAVRCS 57
DB 1 MPDAEQLPDQWKEKTSRSTGMSYLLNMYTKESQWDQTEPAKTKGGSGAGGDPDVHVC 60
QY 58 HLLVKSQRRRPSWQEKITRKEALELINGYIQIKSGEDFSLASQFSDCSSAK 117
DB 61 LLLLVKXGSRPSSWRANITRKEAQLLEIVRNKIVQOEATFELARISYDCSSAK 120
QY 118 ARGDLGAFSRQMQKPPEDAFALRTGEMSGPVFTDGGIHLIR 161
DB 121 RGDLPFSGQMQKPPEDAFALRTGEMSGPVFTDGGIHLIR 164

RESULT 4
T41093
peptidyl-prolyl cis-trans isomerase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Aug-2000
C;Accession: T41093
R;Purnelle, B.; Goffeau, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, September 1998
A;Reference number: Z21964
A;Accession: T41093
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-175 <PUR>
A;Cross-references: EMBL:AL031535; PIDN:CAA20742.1; GSPDB:GN00068; SPDB:SPCC16C4.03
A;Experimental source: strain 972h; cosmid cl6C4
C;Genetics:
A;Gene: SPDB:SPCC16C4.03
A;Map position: 3
A;Introns: 134/2
C;Superfamily: yeast ESS1 protein; WW repeat homology
F;4-42/Domain: WW repeat homology <WWR>

Query Match 45.6%; Score 388.5; DB 2; Length 175;
Best Local Similarity 48.5%; Pred. No. 9.8e-29;
Matches 82; Conservative 21; Mismatches 51; Indels 15; Gaps 2;
QY 7 LPPGWEKMRSSGRVYFNHITNASQWRPSSGGK-----NGQGE 52
DB 6 LPPGWEKMRSSGRVYFNHITNASQWRPSSGGK-----NGQGE 52
QY 53 ARVRCSHLLVKSQRRRPSWQEKITRKEALELINGYIQIKSGEDFSLASQFSD 112
DB 66 -KIRASHLLVKSQRRRPSWQEKITRKEALELINGYIQIKSGEDFSLASQFSD 124
QY 113 CSAXARGDLGAFSRQMQKPPEDAFALRTGEMSGPVFTDGGIHLIR 161
DB 125 CSAXARGDLGAFSRQMQKPPEDAFALRTGEMSGPVFTDGGIHLIR 173

RESULT 5
S52764
ESS1 protein - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein J1452; protein YUR017c; Ptf1 protein
C;Species: Saccharomyces cerevisiae
C;Date: 19-May-1995 #sequence_revision 01-Sep-1995 #text_change 29-Sep-1999
A;Accession: S52764; S55207; S57032; S07867; S65929
R;Hani, J.; Stumpf, G.; Domdey, H.
submitted to the EMBL Data Library, March 1995
A;Description: Ptf1 encodes an essential protein in Saccharomyces cerevisiae, which show
A;Reference number: S52762
A;Accession: S52764
A;Molecule type: DNA
A;Residues: 1-170 <HAN>
A;Cross-references: EMBL:X8761; NID:g758283; PIDN:CAA59961.1; PID:g758286
R;de Haan, M.; Smits, P.H.M.; Grivell, L.A.
submitted to the EMBL Data Library, May 1995
A;Reference number: S55183
A;Accession: S55207
A;Molecule type: DNA
A;Residues: 'MLIVLPTNRLRHNDAP', 1-170 <DBH>
A;Cross-references: EMBL:X8761; NID:g854567; PIDN:CAA60941.1; PID:g854592
R;de Haan, M.; Grivell, L.A.; Smits, P.H.M.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S56771
A;Accession: S57032
A;Molecule type: DNA
A;Residues: 'MLIVLPTNRLRHNDAP', 1-170 <ZAG>
A;Cross-references: EMBL:Z49517; NID:g1015651; PIDN:CAA89541.1; PID:g1015652; MIPS:YUR01
R;Hanes, S.D.; Shank, P.R.; Bostian, K.A.
Yeast 5, 55-72, 1989
A;Title: Sequence and mutational analysis of ESS1, a gene essential for growth in Saccha
A;Reference number: S07867; MUID:89189095; PMID:2648698
A;Accession: S07867
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 'MLIVLPTNRLRHNDAP', 1-7, 'S', '9-16, 'A', 18-126, 'AASA', 133, 'S', 135-136, 'ARC',
A;Experimental source: strain DBY864
R;Hani, J.; Stumpf, G.; Domdey, H.
FEBS Lett. 365, 198-202, 1995
A;Title: Ptf1 encodes an essential protein in Saccharomyces cerevisiae, which shows stro

A;Reference number: S65927; MUID:95300974; PMID:7781779
A;Accession: S65929
A;Molecule type: DNA
A;Residues: 1-170 <HAF>
A;Cross-references: EMBL:X85972; NID:g758283; PIDN:CAA59961.1; PID:g758286
C;Genetics:
A;Gene: SGD:ESS1; PTF1
A;Cross-references: SGD:S0003778; MIPS:YUR017C
A;Map position: 10R
C;Superfamily: yeast ESS1 protein; WW repeat homology
F;9-47/Domain: WW repeat homology <WW1>
Query Match 41.8%; Score 356.5; DB 2; Length 170;
Best Local Similarity 46.3%; Pred. No. 8.7e-26;
Matches 76; Conservative 24; Mismatches 61; Indels 3; Gaps 2;
QY 1 MADEEKLPPGWEKPMSSGRVYFNHITNASQWERPSG-NSSGGKNGQGEPAFVRCSH 59
Db 5 VASRTGLPTWTVRYSKKREYFNPETKHSQWEEPEGTNKDQLKHLRDPVVRCLH 64
QY 60 LLVGHSSRRPSSWRQEKITRTKEALELINGYIQKI--KSGEEDFESLASQFSDCSSAK 117
Db 65 ILIKHDSRRPASHRSNITISKODATDELKTLIRLDDSKTNSFALAKERSDCSSYK 124
QY 118 ARGDLGAFSGQMKPPEDASFALRTGEMSGPVFTDSGIHILR 161
Db 125 RGGDLGFWGREGMOPSFEDAAFLQKVGESVDIVSGSGVHVYKR 168
RESULT 6
B84559
probable peptidyl-prolyl cis-trans isomerase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
A;Accession: B84559
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-119 <STO>
A;Cross-references: GB:AF002093; NID:g4406814; PIDN:AAD20122.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g18040
A;Map position: 2
Query Match 35.3%; Score 301; DB 2; Length 119;
Best Local Similarity 51.3%; Pred. No. 8e-21;
Matches 58; Conservative 25; Mismatches 26; Indels 4; Gaps 1;
QY 54 RVRCSHLVKSQSRPSSWRQEK----ITRTKEALELINGYIQKIKSGEEDFESIASQ 109
Db 6 QVKASHILIKHGSRKRSKWDPEGKIILTTTREAQVQLKSIREDIVSGKANFEVATR 65
QY 110 FSDCSSAKARGDLGAFSGQMKPPEDASFALRTGEMSGPVFTDSGIHILRT 162
Db 66 VSDCSSAKRGDGLGSGFGQMKPFEEATYALKVGDISDVIDTDSGVHILKRT 118
RESULT 7
S15269
Post-translocation molecular chaperone prsa - Bacillus subtilis
N;Alternate names: 33K lipoprotein prsa
C;Species: Bacillus subtilis
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 15-Oct-1999
A;Accession: S15269; I40003; H69682
R;Kontinen, V.P.; Saris, P.; Sarvas, M.
Mol. Microbiol. 5, 1273-1283, 1991
A;Title: A gene (prsa) of Bacillus subtilis involved in a novel, late stage of protein e

A;Reference number: S15269; MUID:92065824; PMID:1956302
A;Accession: S15269
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-292 <KON>
A;Cross-references: EMBL:X57271; NID:g39781; PIDN:CAA40543.1; PID:g39782
R;Smith, H.; de Jong, A.; Bron, S.; Venema, G.
Gene 70, 351-361, 1988
A;Title: Characterization of signal-sequence-coding regions selected from the Bacillus
A;Reference number: I39994; MUID:89108019; PMID:3145906
A;Accession: I40003
A;Status: preliminary; translated from GB/EMBL/DDBU
A;Molecule type: DNA
A;Residues: 'MKIGR', 4, 'LL', 7-8, 'VILS', 13, 'VFSITYLO', 22, 181-195, 'T', 197-210, 'LE', 279-280,
A;Cross-references: GB:M22909; NID:g143687; PIDN:AAA22825.1; PID:g143688
R;Kunst, P.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chc
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galle
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koester, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Serot
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, I
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: H69682
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-292 <KUN>
A;Cross-references: GB:g99109; GB:AL009126; NID:g2633260; PIDN:CAB12835.1; PID:ell82997,
A;Experimental source: strain 168
C;Genetics:
A;Gene: prsa
Query Match 21.2%; Score 180.5; DB 2; Length 292;
Best Local Similarity 39.7%; Pred. No. 3.5e-09;
Matches 46; Conservative 17; Mismatches 30; Indels 23; Gaps 4;
QY 49 QGEPARVRCSHLVKHSQSRPSSWRQEKITRTKEALELINGYIQKIKSGEEDFESIAS 108
Db 131 EGLKGKIRASHILVADK-----KTAEEV-----EKKLKGKX-FEDLAK 168
QY 109 QFSDCSSAKARGDLGAFSR-GQMKPFEDASFALRTGEMSGPVFTDSGIHILRT 163
Db 169 EYSTDSSASKSGDGLGWFPAKEGQMDFTSKAFLKGTGEVDPVKTQYGIHIKTE 224
RESULT 8
A96934
peptidyl-prolyl cis-trans isomerase [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
A;Accession: A96934
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: A96934
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-247 <KUR>
A;Cross-references: GB:AF001437; PIDN:AAK78260.1; PID:g15023118; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC0279

E82712
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C/Accession: E82712
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A/Reference number: A89515; MUID:20365717; PMID:10910347
A/Note: for a complete list of authors see reference number A59328 below
A/Accession: E82712
A>Status: preliminary
A/Molecule type: DNA
A/Residues: 1-655 <SIM>
A/Cross-references: GB:AEO03953; GB:AEO03849; NID:g9106156; PIDN:AAF84001.1.; GSPDB:GN00.1
A/Experimental source: strain 9a5c
R/Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Laig
Briotes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, I
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Proh
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Kriesger, J.E.; Kuramse, E.E.; Laig
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, I
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, E.C.; Miyaki, C.Y.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sasawal
A:Authors da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; ;
A/Reference number: A59328
A/Contents: annotation
C/Genetics:
A/Gene: XF1191

Query Match 17.3%; Score 147; DB 2; Length 655;
Best Local Similarity 28.5%; Pred. No. 1.2e-05;
Matches 39; Conservative 23; Mismatches 57; Indels 18; Gaps 3;

QY 30 NASOWERPQNSS-----SGKXNGQGEPARVCRSHLLVKHSORRPSSNRQEKITRTKE 83
Db 257 DAALHPRPTPEAVLRKYDAQQRGEAQNEQRKAHTLITAGADAASQKVAAEAKALIVE 316
QY 84 EALEELINGVIQTIKSGEEDFESLAS-QFDCSSAKARGDILGAFSRGOMOKFFEDASFALR 142
Db 317 EA-----RKPGVDFAALARINSDPGSKDXGDGLGWVGGMVRFEDVLPAFK 365
QY 143 TGEMSGGVFTDSGIHII 159
Db 366 VGEVVGPBKTEFGNHVI 382

RESULT 15
A81794
probable rotamase NMA2206 [imported] - Neisseria meningitidis (strain Z2491 serogroup A)
C/Species: Neisseria meningitidis
C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C/Accession: A81794
R/Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; More
Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream
Nature 404, 502-506, 2000
A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A/Reference number: A81775; MUID:2022555; PMID:10761919
A/Accession: A81794
A>Status: preliminary
A/Molecule type: DNA
A/Residues: 1-7347 <PAR>
A/Cross-references: GB:AL162758; GB:AL157959; NID:g7380672; PIDN:CAB85417.1; PID:g73808
A/Experimental source: serogroup A, strain Z2491
C/Genetics:
A/Gene: NMA2206

Query Match 17.0%; Score 145; DB 2; Length 347;
Best Local Similarity 37.1%; Pred. No. 8.3e-06;
Matches 43; Conservative 15; Mismatches 44; Indels 14; Gaps 4;

QY	49	QGEPAR-VRCSHLLVKHSOSRRPSSWRQEKITRTKEEALELINGYIQIKSGEEDPESLA	107
Db	201	EGAPLQYRAQHILIK-ADSENAAVGAESTIRKIYGEA-----RSG-TDFSSLA	247
QY	108	SQFSDCSSAKARGDLGAFSRGQMKPFEDASPALRTGEMSGPVFTDSGTHILRTE	163
Db	248	ROYSQDASAGNGDGLGWFADGVVPAPFAFEAVHALKPGQVGAPVRTQFGWHIIKLE	303

Search completed: June 28, 2004, 18:39:59
Job time : 45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2004, 16:36:51 ; Search time 26 Seconds
(without alignments)
326.440 Million cell updates/sec

Title: AAC50492

Perfect score: 852

Sequence: 1 MADEKLPFGWEKMRSSG.....GEMSGPVFTSDGIHILRTE 163

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	852	100.0	163	1 PIN1 HUMAN	Q13526 homo sapien
2	814	95.5	165	1 PIN1 MOUSE	Q9qur7 mus musculus
3	456.5	53.6	166	1 DOD DROME	P54353 drosophila
4	428.5	50.3	182	1 SSF1 NEUCR	O60045 neurospora
5	388.5	45.6	175	1 PIN1 SCHPO	C74448 schizosacch
6	356.5	41.8	190	1 ESS1 YEAST	P22696 saccharomyc
7	301	35.3	119	1 PIN1 ARATH	Q9A142 arabidopsis
8	297	34.9	100	1 PIN1 HUMAN	O15428 homo sapien
9	296.5	34.8	121	1 PIN1 MALDO	Q94500 malus domes
10	283	33.2	118	1 PIN1 DIGLA	Q91ek8 digitalis l
11	180.5	21.2	292	1 PRSA_BACSU	P24327 bacillus su
12	152	17.8	131	1 PIN4 HUMAN	Q9Y237 homo sapien
13	152	17.8	131	1 PIN4 MOUSE	Q9CWW6 mus musculus
14	149	17.5	273	1 CBP2 CAMJE	Q46105 campylobact
15	138	16.2	92	1 PPTC ECOLI	P39159 escherichia
16	137	16.1	92	1 PPTC SALTU	Q916s3 salmonella
17	118	13.8	421	1 GAS7 MOUSE	Q60780 mus musculus
18	118	13.8	422	1 GAS7 RAT	O55148 rattus norv
19	116	13.6	428	1 SURA ECOLI	P21202 escherichia
20	116	13.6	748	1 SUF2 HUMAN	Q9HAU4 homo sapien
21	115.5	13.6	864	1 ITCH MOUSE	Q8C863 mus musculus
22	115.5	13.6	903	1 ITCH HUMAN	Q96302 homo sapien
23	115	13.5	412	1 GAS7 HUMAN	O60861 homo sapien
24	114.5	13.4	887	1 NED4 MOUSE	P46935 mus musculus
25	114	13.4	870	1 WWP2 HUMAN	O40308 homo sapien
26	114	13.4	870	1 WWP2 MOUSE	Q9DBH0 mus musculus
27	112.5	13.2	266	1 NIFM KLEPN	P08534 klebsiella
28	110.5	13.0	293	1 NIFM AZOCH	P23119 azotobacter
29	109	12.8	918	1 WWP1 MOUSE	Q8BZ23 mus musculus
30	108.5	12.7	922	1 WWP1 HUMAN	Q9H0M0 homo sapien
31	107.5	12.6	887	1 NED4 RAT	Q62940 rattus norv
32	107	12.6	299	1 Y175 HELPJ	Q9ZMQ7 helicobacte
33	107	12.6	299	1 Y175_HELPY	P56112 helicobacte

RESULT 1

PIN1 HUMAN
ID PIN1 HUMAN STANDARD; PRT; 163 AA.
AC Q13526;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, last sequence update)
DT 10-OCT-2003 (Rel. 42, last annotation update)
DE Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (BC 5.2.1.8)
DE (Rotamase Pin1) (PPIase Pin1).
GN PIN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96195064; PubMed=8606777;
RA Lu K.P., Hanes S.D., Hunter T.;
RT "A human peptidyl-prolyl isomerase essential for regulation of
RT mitosis."
RL Nature 380:544-547 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RE MEDLINE=22388257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Kleusner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buotow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.35 ÅNGSTRÖMS).
RE MEDLINE=97344079; PubMed=9200606;
RA Ranganathan R., Lu K.P., Hunter T., Noel J.P.;
RT "Structural and functional analysis of the mitotic rotamase Pin1
RT suggests substrate recognition is phosphorylation dependent."
RL Cell 89:875-886 (1997).
CC -!- FUNCTION: Essential PPIase that regulates mitosis presumably by
CC interacting with NIMA and attenuating its mitosis-promoting
CC activity. Displays a preference for an acidic residue N-terminal
CC to the isomerized proline bond. Catalyzing pSer/Thr-Pro cis/trans

ALIGNMENTS

34 107 12.6 622 1 PPID_HAEIN
35 104.5 12.3 577 1 BAG3_MOUSE
36 104 12.2 282 1 PLP_RICCN
37 103 12.1 1000 1 NED4_HUMAN
38 102 12.0 282 1 PLP_RICPR
39 102 12.0 292 1 NIFM_AZOVI
40 101.5 11.9 583 1 PR40_YEAST
41 101 11.9 767 1 PUB1_SCHPO
42 100 11.7 704 1 CT67_HUMAN
43 100 11.7 706 1 CT67_MOUSE
44 99.5 11.7 809 1 RSP5_YEAST
45 99 11.6 472 1 YAP1_MOUSE

P44092 haemophilus
Q9jlv1 mus musculus
Q29h91 rickettsia
P46934 homo sapien
Q9ZCX6 rickettsia
P14890 azotobacter
P33203 saccharomyc
Q32462 schizosacch
Q9H433 homo sapien
P59114 mus musculus
P39940 saccharomyc
P46938 mus musculus

```

CC isomerizations.
CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the ppiC/parvulin rotamase family.
CC -!- SIMILARITY: Contains 1 WW domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U49070; AAC50492.1; -.
CC EMBL; BC002899; AAH02899.1; -.
CC PIR; S68520; S68520.
CC PDB; 1PIN; 25-NOV-98.
CC PDB; 1F8A; 23-AUG-00.
CC PDB; 118G; 18-JUL-01.
CC PDB; 118H; 18-JUL-01.
CC PDB; 118H; 18-JUL-01.
CC Genew; HGNC:8988; PIN1.
CC GK; Q13526; -.
CC MIM; 601052; -.
CC GO; GO:0005634; C:nucleus; TAS.
CC GO; GO:0005515; P:protein binding; TAS.
CC GO; GO:0007088; P:regulation of mitosis; TAS.
CC InterPro; IPR000297; Rotamase.
CC InterPro; IPR001202; WW_Rsp5_WWP.
CC Pfam; PF00639; Rotamase; 1.
CC Pfam; PF00397; WW; 1.
CC SMART; SM00456; WW; 1.
CC PROSITE; PS01096; PPI_C_PPIASE_1; 1.
CC PROSITE; PS01098; PPI_C_PPIASE_2; 1.
CC PROSITE; PS01159; WW_DOMAIN_1; 1.
CC PROSITE; PS01159; WW_DOMAIN_2; 1.
CC PROSITE; PS01159; WW_DOMAIN_3; 1.
CC Isomerase; Rotamase; Nuclear protein; Cell cycle; 3D-structure.
CC KW
CC FT DOMAIN 5 39
CC FT TURN 52 163
CC FT STRAND 11 15
CC FT TURN 17 18
CC FT STRAND 22 26
CC FT TURN 27 29
CC FT STRAND 32 33
CC FT TURN 55 62
CC FT STRAND 72 72
CC FT TURN 73 74
CC FT STRAND 75 75
CC FT TURN 82 98
CC FT HELIX 99 99
CC FT HELIX 103 110
CC FT HELIX 114 118
CC FT TURN 119 120
CC FT STRAND 121 125
CC FT TURN 127 128
CC FT HELIX 132 140
CC FT TURN 143 144
CC FT STRAND 146 146
CC FT STRAND 150 151
CC FT STRAND 156 161
CC SEQUENCE 163 AA; 18243 MW; 35391AF40B7D1E13 CRC64;

Query Match 100.0%; Score 852; DB 1; Length 163;
Best Local Similarity 100.0%; Pred. No. 3.9e-72;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60
Db 1 MADEKLPPGWEKMRSSGRVYFNHITNASQWERPSGNSGGKNGQGEPAVRCSHL 60

```

```

QY 61 LVKHSQSRPSSRWQEKITRTKEALELNGYIQIKSGEDEFESLASQFSCSSAKARG 120
Db 61 LVKHSQSRPSSRWQEKITRTKEALELNGYIQIKSGEDEFESLASQFSCSSAKARG 120
QY 121 DLGAFSRGQMQKPFEDASFALETGEMSGPVFTDSGHIILRT 163
Db 121 DLGAFSRGQMQKPFEDASFALETGEMSGPVFTDSGHIILRT 163

RESULT 2
PIN1_MOUSE
ID PIN1_MOUSE STANDARD; PRT; 165 AA.
AC Q9OUR7;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1 (EC 5.2.1.8)
DE (Rotamase Pini) (PPIase Pini).
GN PIN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20070807; PubMed=10600477;
RA Fujimori F., Takahashi K., Uchida C., Uchida T.;
RT "Mice lacking Pini develop normally, but are defective in entering
RL cell cycle from G0 arrest.";
RL Biochem. Biophys. Res. Commun. 265:658-663(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Embryo and Kidney;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.P.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayaishizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: Essential PPIase that regulates mitosis presumably by
CC interacting with NIMA and attenuating its mitosis-promoting
CC activity. Displays a preference for an acidic residue N-terminal
CC to the isomerized proline bond. Catalyzing pSer/Thr-Pro cis/trans
CC isomerizations (By similarity).
CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: Belongs to the ppiC/parvulin rotamase family.
CC -!- SIMILARITY: Contains 1 WW domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC  DR EMBL; AB009691; BAA87037.1; -
CC  DR EMBL; AB009692; BAA87038.1; -
CC  DR EMBL; AK003369; BAA22743.1; -
CC  DR EMBL; AK002665; BAA22270.1; -
CC  DR PIR; JCT136; JCT136.
CC  DR HSSP; Q13526; 1PIN.
CC  DR MGD; MGI:1346036; Pin1.
CC  DR GO; GO:0042127; P;regulation of cell proliferation; IMP.
CC  DR InterPro; IPR000297; Rotamase.
CC  DR InterPro; IPR001202; WW_Rsp5_WWP.
CC  DR Pfam; PF00639; Rotamase; 1.
CC  DR Pfam; PF00397; WW; 1.
CC  DR SMART; SM00456; WW; 1.
CC  DR PROSITE; PS01096; PPIC_PPIASE_1; 1.
CC  DR PROSITE; PS01098; PPIC_PPIASE_2; 1.
CC  DR PROSITE; PS01159; WW_DOMAIN_1; 1.
CC  DR PROSITE; PS01020; WW_DOMAIN_2; 1.
CC  KW Isomerase; Rotamase; Nuclear protein; Cell cycle.
CC  FT DOMAIN 5 39 WW.
CC  FT DOMAIN 54 165 PPIC.
CC  SQ SEQUENCE 165 AA; 18370 MW; 189E95F009176B1F CRC64;

Query Match
Best Local Similarity 95.5%; Score 814; DB 1; Length 165;
Matches 157; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 1 MADEKLPGWWEKMRSSGRVYFNHITNASQWERPSGNSSSGG--KNGQGEPAKVRCS 58
Db 1 MADEKLPGWWEKMRSSGRVYFNHITNASQWERPSGNSSSGG--KNGQGEPAKVRCS 60
QY 59 HLLVHQSRRPSSWRQEKITTKBEALELINGYIQKINSBEDFESLASQFSDCSSAKA 118
Db 61 HLLVHQSRRPSSWRQEKITTKBEALELINGYIQKINSBEDFESLASQFSDCSSAKA 120
QY 119 RGLDFAFRSGQKQPFEDASFAIRTGEMSGPVFTDSGTHILRTE 163
Db 121 RGLDGFRRSGQKQPFEDASFAIRTGEMSGPVFTDSGTHILRTE 165

RESULT 3
DOD DROME STANDARD; PRT; 166 AA.
AC P5A355; O61344; Q9VRH1;
DT 01-OCT-1996 (Rel. 34; Created)
DT 16-OCT-2001 (Rel. 46; Last sequence update)
DT 10-OCT-2003 (Rel. 42; Last annotation update)
DE Dodo protein.
GN DOD OR CGI7051.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R; TISSUE=Larva, and Pupae;
RX MEDLINE=96133954; PubMed=8552658;
RA Maleszka R., Hanes S.D., Hackett R.L., de Couet H.G., Miklos G.L.G.;
RT "The Drosophila melanogaster dodo (dod) gene, conserved in humans, is
RT functionally interchangeable with the E5S1 cell division gene of
RT Saccharomyces cerevisiae."
RL Proc. Natl. Acad. Sci. U.S.A. 93:447-451(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=98188272; PubMed=9520435;
RA Maleszka R., de Couet H.G., Miklos G.L.G.;
RT "Data transferability from model organisms to human beings: insights
RT from the functional genomics of the flightless region of Drosophila.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:3731-3736(1998).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=Berkeley;
RC MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhargava M., Boldt J.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA deFabos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Ketchum K.A.,
RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murthy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner K., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
CC -!- SIMILARITY: Belongs to the ppiC/parvulin rotamase family.
CC -!- SIMILARITY: Contains 1 WW domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U35140; AAC46958.1; -
CC EMBL; AF017777; AAC28408.1; -
CC EMBL; AE003568; AAF50829.1; -
CC PIR; T08426; T08426.
CC HSSP; Q13526; 1PIN.
CC FlyBase; FBgn015379; dod.
CC InterPro; IPR000297; Rotamase.
CC InterPro; IPR001202; WW_Rsp5_WWP.
CC Pfam; PF00639; Rotamase; 1.
CC Pfam; PF00397; WW; 1.
CC SMART; SM00456; WW; 1.
CC PROSITE; PS01096; PPIC_PPIASE_1; 1.
CC PROSITE; PS01098; PPIC_PPIASE_2; 1.
CC PROSITE; PS01159; WW_DOMAIN_1; 1.
CC PROSITE; PS01020; WW_DOMAIN_2; 1.
CC Isomerase; Rotamase.
CC DOMAIN 5 39 WW.
CC DOMAIN 55 166 PPIC.
CC FT CONFLICT 44 A -> T (IN REF. 2).
CC SQ SEQUENCE 166 AA; 18376 MW; 3B4306FA930E7259 CRC64;

Query Match
53.6%; Score 456.5; DB 1; Length 166;

```

Best Local Similarity 56.7%; Pred. No. 2e-35;
Matches 93; Conservative 16; Mismatches 52; Indels 3; Gaps 2;

QY 1 MADEKLPGEKMSRSGRVYFNHTNASHQWERS--GNSSGGKNGQE-PARVRC 57
DB 1 MPDAEQPDGVEKRTSRSTGMSYLYNMYTKESQWDPTEPAKAGGSGAGGDADEVHC 60
QY 58 SHLLVKHSQSRPPSRWROEKTRTKEAELELINGYQIKKSGEEDFESLASQFSDCSAK 117
DB 61 LHLVVKHSGSRPPSRWRANITRTKEAQLLEVTNRKIVQEQATFDELARSYSDCSAK 120
QY 118 ARGDLGAFSRGQKPFEDAFALRTGEMSGPVFTDGIHILR 161
DB 121 RGGDLGKFGRCQMAAFEDAAFKLVNQLSGIVDSGLHILR 164

RESULT 4
SSP1_NEUCR
ID SSP1_NEUCR STANDARD; PRT; 182 AA.
AC O60045;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase sspl (EC 5.2.1.8) (PPIase sspl).
GN SSP1.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=99041963; PubMed=9822668;
RA Kops O.; Eckerskorn C.; Hottenrott S.; Fischer G.; Mi H.;
RA Tropisch M.;
RT "Sspl, a site specific parvulin homolog from N.crassa active in
RT protein folding";
RL J. Biol. Chem. 273:31971-31976 (1998).
CC -!- FUNCTION: Site-specific PPIase with respect to the amino acid N-
CC terminal to the proline residue. Peptides with glutamate,
CC phosphoserine, or phosphothreonine in the -1 position are the best
CC substrates. It is not only able to isomerize small peptides but
CC is also active in protein folding.
CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -!- SIMILARITY: Contains 1 WW domain.
CC -!- SIMILARITY: Belongs to the ppiC/parvulin rotamase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AJ004603; CAA06818.1; -
CC DR HSP; Q13526; IPIN.
CC DR InterPro: IPR000297; Rotamase.
CC DR InterPro: IPR001202; WW_Rsp5_WWP.
CC DR Pfam: PF00639; Rotamase; 1.
CC DR Pfam: PF00397; WW; 1.
CC DR SMART: SM00456; WW; 1.
CC DR PROSITE: PS0198; PPIase 2; 1.
CC DR PROSITE: PS0159; WW_DOMAIN 1; 1.
CC DR PROSITE: PS0020; WW_DOMAIN 2; 1.
CC KW Isomerase; Rotamase.
FT DOMAIN 7 41 WW
FT DOMAIN 60 71 GLN-RICH.
FT DOMAIN 71 182 PPI.
SQ SEQUENCE 182 AA; 20673 MW; ACBD0497DFF7339 CRC64;

Query Match 50.3%; Score 428.5; DB 1; Length 182;
Best Local Similarity 50.8%; Pred. No. 8.7e-33;

Matches 90; Conservative 20; Mismatches 50; Indels 17; Gaps 2;

QY 4 BEKLPPGWEKMSRSGRVYFNHTNASHQWERSG-----NSSSGKN 47
DB 6 ETGLPEDVEVHSGKNLFFYFNSATKTSRWEPPSGTDVVKLTMYAKYHSFTSQOQQQ 65
QY 48 GQGP-ARVRCSHLLVKHSQSRPPSRWROEKTRTKEAELELINGYQIKKSGEEDFESL 106
DB 66 QQQQPOGKIRCAHLLVKHNSRRPSSRWESEITTKAEALTTQGFQFRKNGSISLGL 125
QY 107 ASQFSDCSAKARGDLGAFSRGQKPFEDAFALRTGEMSGPVFTDGIHILRTE 163
DB 126 ALTESDCSSARKRGDLGFGFGDMQKBEFEDAAFAFKPEISDVDTASGLHLIERLE 182

RESULT 5
PIN1_SCHPO
ID PIN1_SCHPO STANDARD; PRT; 175 AA.
AC O7448;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase pini (EC 5.2.1.8).
GN PIN1 OR SPCL164.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetiales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V.; Gwilliam R.; Rajandream M.A.; Lyne M.; Lyne R.; Stewart A.;
RA Scouras J.; Beat N.; Hayles J.; Baker S.; Basham D.; Bowman S.;
RA Brooks K.; Brown D.; Brown S.; Chillingworth T.; Churcher C.M.;
RA Collins M.; Connor R.; Cronin A.; Davis P.; Feltwell T.; Fraser A.;
RA Gentles S.; Goble A.; Hamlin N.; Harris D.; Hidalgo J.; Hodgson G.;
RA Holroyd S.; Hornsby T.; Howarth S.; Huckle E.J.; Hunt S.; Jagels K.;
RA James K.; Jones L.; Jones M.; Leather S.; McDonald S.; McLean J.;
RA Meeney P.; Moule S.; Mungall K.; Murphy L.; Niblett D.; Odeh C.;
RA Oliver K.; O'Neill S.; Pearson D.; Quail M.A.; Rabinowitsch E.;
RA Rutherford K.; Rutter S.; Saunders D.; Seeger K.; Sharp S.;
RA Skelton J.; Simmonds M.; Squares R.; Stevens K.;
RA Taylor K.; Taylor R.G.; Tivey A.; Walsh S.V.; Warren T.; Whitehead S.;
RA Woodward J.; Volckaert G.; Aert R.; Robben J.; Grynoprez B.;
RA Welljens I.; Vanstreels E.; Rieger M.; Schaefer M.; Mueller-Auer S.;
RA Gabel C.; Fuchs M.; Fritz C.; Holzer E.; Moestl D.; Hilbert H.;
RA Borzym K.; Langer I.; Beck A.; Leinhardt R.; Pohl T.M.;
RA Eger P.; Zimmermann W.; Wedler H.; Reinhardt R.; Purnelle B.;
RA Coffeau A.; Cadieu E.; Dreano S.; Gloux S.; Lelaure V.; Mottier S.;
RA Galibert F.; Aves S.J.; Xiang Z.; Hunt C.; Moore K.; Hurst S.M.;
RA Lucas M.; Rochet M.; Gallardin C.; Tallada V.A.; Garzon A.; Thode G.;
RA Daga R.R.; Cruzado L.; Jimenez J.; Sanchez M.; del Rey F.; Benito J.;
RA Dominguez A.; Revuelta J.L.; Moreno S.; Armstrong J.; Forsburg S.L.;
RA Cerretti L.; Lowe T.; McCombie W.R.; Paulsen I.; Potashkin J.;
RA Shpakovski G.V.; Ussery D.; Barrell B.G.; Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe";
RL Nature 415:871-880 (2002).
RN [2]
RP FUNCTION, AND SUBCELLULAR LOCATION.
RX MEDLINE=21564328; PubMed=11707530;
RA Huang H.-K.; Forsburg S.L.; John U.P.; O'Connell M.J.; Hunter T.;
RT "Isolation and characterization of the Pini/Esslp homologue in
RT Schizosaccharomyces pombe";
RL J. Cell Sci. 114:3779-3788 (2001).
CC -!- FUNCTION: Has a role in the G1/S stage transition of mitosis
CC where it is involved in the dephosphorylation of cdc25 and weel.
CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: Belongs to the ppiC/parvulin rotamase family.
CC -!- SIMILARITY: Contains 1 WW domain.

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL031535; CAA20742.1; --
DR PIR; T41093; T41093.
DR HSSP; Q13526; SPCC16C4.03; --.
DR GeneDB_SPombe; SPCC16C4.03; --.
DR InterPro; IPR000297; Rotamase.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00639; Rotamase; 1.
DR Pfam; PF00397; WW; 1.
DR SMART; SM00456; WW; 1.
DR PROSITE; PS01096; PPIC_PPIASE_1; FALSE_NEG.
DR PROSITE; PS01198; PPIC_PPIASE_2; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; FALSE_NEG.
DR PROSITE; PS00020; WW_DOMAIN_2; 1.
KW Isomerase; Rotamase; Nuclear protein; Mitosis.
FT DOMAIN 4 38
FT DOMAIN 64 175
FT SEQUENCE 175 AA; 19773 MW; ABA637835471BD25 CRC64;
Query Match 45.6%; Score 388.5; DB 1; Length 175;
Best Local Similarity 48.5%; Pred. No. 4.3e-29;
Matches 82; Conservative 21; Mismatches 55; Indels 15; Gaps 2;
QY 7 LPQWEKMSRSSGRVYFNHITNASQWERPSGNSGGK-----NGQGEF 52
Db 6 LPKFWIKISRSRNPFFNTTHESLWEPDAATDMAALKXFIANELQESVTPTEANSF 65
QY 53 ARVCSSHLVVKHSRPPSSNRQKTRTKKEALELINGYIOKTKSGEEDFESLASFSD 112
Db 66 -KIPASHLLVGRSSRPPSSKKEEHTKSKKEARKLAHEYBQLLKSGSVSMHDLAMKESD 124
QY 113 CSSNAKRGDIAFGSRGQKQPFEDASFAIRTGEMSGPVFTDSGTHILR 161
Db 125 CSSARRGGELGEGFGRDEMQRPFEDAAFAALKGEISGVVETSSGHHILQR 173
RESULT 6
ESS1_YEAST STANDARD; PRT; 190 AA.
AC P22636;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ESS1 protein (Processing/termination factor 1).
GN ESS1 OR PTF1 OR YJR017C OR J1452.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89189095; PubMed=2648698;
RA Hanes S.D., Shank P.R., Bostian K.A.;
RT "Sequence and mutational analysis of ESS1, a gene essential for
RT growth in Saccharomyces cerevisiae.";
RL Yeast 5:55-72(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=DH484;
RX MEDLINE=95300974; PubMed=7781779;
RA Hani J., Stumpf G., Domdey H.;
RT "PTF1 encodes an essential protein in Saccharomyces cerevisiae, which
RT shows strong homology with a new putative family of PPIases.";
RL FEBS Lett. 365:198-202(1995).
RN [3]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=S288c/F1679;
RA de Haan M., Smits P.H.M., Grivell L.A.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: Essential for growth in Saccharomyces cerevisiae. May be
CC involved in cytokinesis or in cell separation.
CC !- SIMILARITY: Contains 1 WW domain.
CC !- SIMILARITY: Belongs to the pp1C/parvulin rotamase family.
CC !- CAUTION: It is uncertain whether Met-1 or Met-21 is the initiator.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X85972; CAA59961.1; ALT_INIT.
DR EMBL; X87611; CAA60941.1; --.
DR EMBL; Z49517; CAA89541.1; --.
DR PIR; S52764; S52764.
DR HSSP; Q13526; 1PIN.
DR GerMOnline; 141854; --.
DR SGD; S0003778; BSS1.
DR GO; GO:0005634; C:nucleus; IPI.
DR GO; GO:0003755; F:peptidyl-prolyl cis-trans isomerase activity; IDA.
DR GO; GO:0006397; P:mRNA processing; IMP.
DR GO; GO:0042326; P:negative regulation of phosphorylation; IDA.
DR InterPro; IPR000297; Rotamase.
DR InterPro; IPR001202; WW_Rsp5_WWP.
DR Pfam; PF00639; Rotamase; 1.
DR Pfam; PF00397; WW; 1.
DR SMART; SM00456; WW; 1.
DR PROSITE; PS01096; PPIC_PPIASE_1; 1.
DR PROSITE; PS01198; PPIC_PPIASE_2; 1.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS00020; WW_DOMAIN_2; 1.
KW Isomerase; Rotamase.
FT DOMAIN 29 62
FT DOMAIN 77 190
FT CONFLICT 28 28 R -> S (IN REF. 1).
FT CONFLICT 37 37 V -> A (IN REF. 1).
FT CONFLICT 147 190 GDLGWFGRGEMQSFEDAAFAALKVGEVSDIVESGSGVHVIK
FT REV -> AASAGSGEARCSLALKTLFSSSSAR (IN
FT REF. 1).
SQ SEQUENCE 190 AA; 21733 MW; 93E449E3D7B4D989 CRC64;
Query Match 41.8%; Score 356.5; DB 1; Length 190;
Best Local Similarity 46.3%; Pred. No. 4.4e-26;
Matches 75; Conservative 24; Mismatches 61; Indels 3; Gaps 2;
QY 1 MADEKLPPGWEKMRSGRVRVYFNHITNASQWERPSG-NSSSGKNGQGEFARVRCSH 59
Db 25 VASRTGLPTWTVRYSKSKREYFFNPETKHSQWEEPEGTKDQLKHLRDHPVRVRLH 84
QY 60 LLVVKHSRPPSSNRQKTRTKKEALELINGYIOKI--KSGEEDFESLASFDCSSAK 117
Db 85 ILIKHKDSRRPASHRSNTISKQDATDELKTLITRLDDDSKTNSEALAKERSDCSSYK 144
QY 118 ARGDLGAFSRGQKQPFEDASFAIRTGEMSGPVFTDSGTHILR 161
Db 145 RGDLGWFGRGEMQSFEDAAFAALKVGEVSDIVESGSGVHVIR 188
RESULT 7
PIN1_ARATH STANDARD; PRT; 119 AA.
AC Q9SL42; Q42334;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase 1 (EC 5.2.1.8) (Rotamase Pin1)

```



```

Db 131 BGLKGKIRASHILVADK-----KTAEEV-----EKLLKGGEK-FEDLAK 168
Qy 109 QFSCSSAKARGDLGAFGR-GOMOKPFEDASFAIRTGEMSGPVFTDSIHILILTE 163
Db 169 EYSTDSASKSGDGLGWFAKEGQMDFTFSKAAFKUKLTGEVDPVTKYGYHIIRKTE 224

RESULT 12
PIN4 HUMAN
ID PIN4 HUMAN STANDARD; PRT; 131 AA.
AC Q9Y237;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4 (EC 5.2.1.8)
DE (Rotamase Pin4) (PPIase Pin4) (Parvulin 14) (Parvulin 14) (Peptidyl-prolyl
DE cis/trans isomerase EPVH) (hPar14).
GN PIN4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=99198957; PubMed=10100858;
RX Uchida T., Fujimori F., Tradler T., Fischer G., Rahfeld J.-U.;
RA "Identification and characterization of a 14 kDa human protein as a
RT novel parvulin-like peptidyl prolyl cis/trans isomerase.";
RL FEBS Lett. 446:278-282(1999).
RN [2]
RN SEQUENCE FROM N.A.
RX MEDLINE=99294627; PubMed=10364457;
RX Rulten S.L., Thorpe J.R., Kay J.E.;
RA "Identification of eukaryotic parvulin homologues: a new subfamily of
RT peptidylprolyl cis-trans isomerases.";
RL Biochem. Biophys. Res. Commun. 259:557-562(1999).
RN [3]
RN SEQUENCE FROM N.A.
RX TISSUE=Urinary bladder;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Shenmen C.M., Schuler G.D.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J.M., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Besak S.A., McEwan P.J., McKernan K.J., Shevchenko Y., Bouffard G.G.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krywinski M.I., Skalska U., Smalhus D.E.,
RA Schnerch A., Schein J.Y., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RL human and mouse cDNA sequences.";
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RN STRUCTURE BY NMR OF 36-131.
RX MEDLINE=20425097; PubMed=10966801;
RX Sekerina E., Rahfeld J.-U., Mueller J., Fanghaenel J., Rascher C.,
RA Fischer G., Bayer P.;
RT "NMR solution structure of hPar14 reveals similarity to the peptidyl
RT prolyl cis/trans isomerase domain of the mitotic regulator hPini But
RT indicates a different functionality of the protein.";
RL J. Mol. Biol. 301:1003-1017(2000).
RN [5]
RN STRUCTURE BY NMR OF 28-131.
RX MEDLINE=21103679; PubMed=11162102;
RX Terada T., Shirouzu M., Fukumori Y., Fujimori F., Ito Y., Kigawa T.,
RA

```

```

RA Yokoyama S., Uchida T.;
RT "Solution structure of the human parvulin-like peptidyl prolyl
RT cis/trans isomerase, hPar14.";
J. Mol. Biol. 305:917-926(2001).
CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC -!- SIMILARITY: Belongs to the pPIC/parvulin rotamase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB009690; BAA82320.1; -
DR EMBL; AF143096; AAD27893.1; -
DR EMBL; BC005234; AAH05234.1; -
DR Genew; HGNC:8992; PIN4.
DR MIM; 300252; -
DR PDB; 1EQ3; 04-APR-01.
DR PDB; 1FJD; 08-AUG-01.
DR GO; GO:0005759; C:mitochondrial matrix; TAS.
DR GO; GO:0006457; P:protein folding; TAS.
DR InterPro; IPR000297; Rotamase.
DR Pfam; PF00639; Rotamase; 1.
DR PROSITE; PS01096; PPI_PPIASE_1; FALSE_NEG.
DR PROSITE; PS01098; PPI_PPIASE_2; 1.
DR Isomerase; Rotamase; 3D-structure.
FT DOMAIN 35 129 PPI_C.
SQ SEQUENCE 131 AA; 12910 MW; 787C15BDB0701258 CRC64;

Query Match 17.8%; Score 152; DB 1; Length 131;
Best Local Similarity 34.0%; Pred. No. 2.6e-07;
Matches 51; Conservative 14; Mismatches 33; Indels 52; Gaps 9;

Qy 37 PSGNSSSG-----GKNGQGEPARVRCSHLLV-KHSQRSPSSWRQE 76
Db 3 PKGSGSGKAGKGAASGSADKKAQPGKGGNAVKR--HILCEKHG----- 49
Qy 77 KITRTKEALELINGYTIKISGEEDPEISLASFDCSSAKARGDLGAFSGOMOKPFED 136
Db 50 KIM-----EAME-----KLKSGMR-FNEVAQYSE-DKARQGDGLGWMTGSGMVGPQE 96
Qy 137 ASFAIRTGEMSGPVFTDS-----GIHIL 160
Db 97 AAFALPVSGMDKPVFTDPPVTKRFGYHIIM 126

RESULT 13
PIN4 MOUSE
ID PIN4 MOUSE STANDARD; PRT; 131 AA.
AC Q9CWW6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4 (EC 5.2.1.8)
DE (Rotamase Pin4) (PPIase Pin4).
GN PIN4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Embryonic stem cells;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

```

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wegner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Bojelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzairelli J., Nombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Hayashizaki Y.,
RA Yashaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohatsu S.,
RT Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- CATALYTIC ACTIVITY: Peptidylproline (omega=180) = peptidylproline
CC (omega=0).
CC -!- SIMILARITY: Belongs to the pp1C/parvulin rotamase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AK010338; BAB26863.1; -.
CC HSPSP; Q9V237; 1803.
CC MGD; MG1.1916963; Pin4.
CC InterPro; IPR000297; Rotamase.
CC Pfam; PF00639; Rotamase; 1.
CC PROSITE; PS01096; PPIC PPIASE 1; FALSE_NEG.
CC PROSITE; PS01096; PPIC PPIASE 2; 1.
CC KW Isomerase; Rotamase.
CC FT DOMAIN 35 129 PPTC.
CC SEQUENCE 131 AA; 13815 MW; 290BDEF72DC69CA6 CRC64;

Query Match 17.8%; Score 152; DB 1; Length 131;
Best Local Similarity 34.0%; Pred. No. 2.6e-07;
Matches 51; Conservative 14; Mismatches 33; Indels 52; Gaps 9;

QY 37 PGNSSSG-----GKNGQEPARVRCSHLLV-KHSQSRPSSWRQE 76
Db 3 PKGSGSGKGGKGAASGSDSADKSGQPKGGNAVKVR--HTLCKHG-----49
QY 77 KITRTKEALELINGVIQIKSGEEDFESLASQFDCSSAKARGDILGAFSRGOMQKPFED 136
Db 50 KIM-----EAME-----KLKSGMR-FSEVATQISE-DRARQGGDLGWMTRGSMVGFQF 96
QY 137 ASFALRTGMSGGVFTDS-----GTHILL 160
Db 97 AAALPVSQMDKPVFTDPVVKTKFGVHIIM 126

RESULT 14
ID CBF2_CAMJF STANDARD; PRT; 273 AA.
AC Q46105;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cell binding factor 2 precursor (Major antigen pib4A).
GN CBF2 OR PEB4A OR CU0596.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 55026 / 81-176;
RX MEDLINE=96099687; PubMed=8525063;

RA Buruoca C., Premeaux C., Pei Z., Tummuru M., Blaser M.J.,
RA Cenatiempo Y., Fauchere J.L.;
RT "Nucleotide sequence and characterization of pib4A encoding an
RT antigenic protein in Campylobacter jejuni.";
RL Res. Microbiol. 146:467-476(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Baeham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC -!- SIMILARITY: Belongs to the pp1C/parvulin rotamase family.
CC STRONG, TO H.FYLORI HP0175.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X84703; CAAS9175.1; -.
CC ENBL; AL139075; CAB75232.1; -.
CC PIR; S52412; S52412.
CC HSPSP; Q13526; 1PIN.
CC InterPro; IPR000297; Rotamase.
CC Pfam; PF00639; Rotamase; 1.
CC PROSITE; PS01096; PPIC PPIASE 1; 1.
CC PROSITE; PS01096; PPIC PPIASE 2; 1.
CC KW Isomerase; Rotamase; Antigen; Signal; Complete proteome.
CC FT SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 22 273 CELL BINDING FACTOR 2.
CC FT DOMAIN 131 228 PPIC.
CC SEQUENCE 273 AA; 30518 MW; 1C014658B8BC4E39 CRC64;

Query Match 17.5%; Score 149; DB 1; Length 273;
Best Local Similarity 36.5%; Pred. No. 1.2e-06;
Matches 42; Conservative 20; Mismatches 31; Indels 22; Gaps 5;

QY 51 EPARVRCSHLLVKSQSRPSSWRQEKITRTKEALELINGVIQIKSGEED--FESLAS 108
Db 130 KPARVQAKHILV-----ATEKEAKDIIN-ELKGLKGLKELDAKFSELAK 171
QY 109 QFS-DCSSAKARGDILGAFSRGOMQKPFEDASFLRTGEM-SGFVTDSCIHIILR 161
Db 172 EKSIDPGSKNGGELGWFEDQSTWVPFTDAAPALKNGTITTPVKTNGYHILK 226

RESULT 15
ID PPIC_ECOLI STANDARD; PRT; 92 AA.
AC P39159;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Peptidyl-prolyl cis-trans isomerase C (EC 5.2.1.8) (PPIase C)
DE (Rotamase C) (parvulin).
GN PPIC OR PARVA OR B3775 OR C4697 OR Z5286 OR ECS4709.
OS Escherichia coli,
OS Escherichia coli O6, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992, 83334;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2004, 14:35:24 ; Search time 3445.82 Seconds
(without alignments)
8614.203 Million cell updates/sec

Title: U49070

Perfect score: 994

Sequence: 1 TGTGCGCCAGCACCTCGAGG.....CCCAATTAACCCAGACCA 994

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *

1: em_estba: *

2: em_esthum: *

3: em_estin: *

4: em_estmu: *

5: em_estov: *

6: em_estpl: *

7: em_estro: *

8: em_hic: *

9: gb_est1: *

10: gb_est2: *

11: gb_hic: *

12: gb_est3: *

13: gb_est4: *

14: gb_est5: *

15: em_estfun: *

16: em_estom: *

17: em_gss_hum: *

18: em_gss_inv: *

19: em_gss_pln: *

20: em_gss_vrt: *

21: em_gss_fun: *

22: em_gss_mam: *

23: em_gss_mus: *

24: em_gss_pro: *

25: em_gss_rtd: *

26: em_gss_phg: *

27: em_gss_vrl: *

28: gb_gss1: *

29: gb_gss2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	895.2	90.1	1035	9	AL526064
C 2	891.4	89.7	988	9	AL526011
C 3	878.8	88.4	950	12	BG744866
4	859	86.4	1076	13	BX436611

5	848.6	85.4	923	12	BI829812
6	817.2	82.2	1201	13	BX463403
C 7	806.2	81.1	1097	9	AL529668
8	805.4	81.0	868	9	AL533491
C 9	799.8	80.5	1039	13	BX436610
10	792.8	79.8	1201	13	BX421727
11	784.2	78.9	864	13	BU557229
12	781	78.6	928	13	BQ678260
13	776.8	78.1	833	12	BG744294
14	775.6	78.0	964	12	BI753931
15	772.2	77.7	874	10	BE797127
16	768.8	77.3	819	12	BI871560
17	762	76.7	837	12	BI597391
18	757.6	76.2	952	13	BQ948365
19	753.6	75.8	1075	13	BX361953
20	750.2	75.5	772	12	BI590834
21	745.4	75.0	921	12	BI596653
C 22	737.4	74.2	797	14	CD364657
23	735	73.9	954	13	BQ068942
24	729.8	73.4	1089	9	AL529669
25	729.2	73.4	854	14	CA454466
26	726	73.0	938	13	BQ430612
27	725.6	73.0	963	13	BU500536
28	724	72.8	919	13	BU182574
29	716.6	72.1	935	14	CA487282
C 30	716	72.0	765	14	CB321634
C 31	716	72.0	772	14	CD365653
C 32	714.8	71.9	817	10	BF570461
33	714	71.8	982	13	BQ956997
34	712.8	71.7	816	12	BG471419
35	700	70.4	712	12	BI489909
C 36	693.4	69.8	767	13	BU683491
37	682.8	68.7	937	9	AU123122
C 38	677.4	68.1	740	14	CD367366
39	676.4	68.0	1200	10	BF314489
40	674.4	67.8	923	10	BF793524
41	671	67.5	728	9	AU123171
42	669.2	67.3	726	10	BE893240
43	669.2	67.3	742	12	BG395172
44	665.8	67.0	630	12	BI856005
45	665.6	67.0	511	14	CA487426

ALIGNMENTS

RESULT 1
AL526064
LOCUS
DEFINITION
AL526064 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
cDNA clone CS0DC015YK14 5-PRIME, mRNA sequence.
1035 bp mRNA linear EST 23-MAY-2003
ACCESSION
AL526064
VERSION
AL526064.2 GI:31063925
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 1035)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 13, 2001 this sequence version replaced gi:12789557.
CONTACT: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9850.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DC015BF07QF1&cluster=9850.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DC015BF07QPL.

FEATURES

source
location/Qualifiers
1..1035
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC015YK14"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 90.1%; Score 895.2; DB 9; Length 1035;
Best Local Similarity 97.9%; Pred. No. 1.4e-170;
Matches 946; Conservative 3; Mismatches 11; Indels 6; Gaps 4;

Qy 29 CGGACGAGGAGAACTGCGCCGCTGGGAGAGCGCATGAGCGCAGCTCAGGCCGAG 88
Db 69 CGGACGAGGAGAACTGCGCCGCTGGGAGAGCGCATGAGCGCAGCTCAGGCCGAG 128
Qy 89 TGTTACTTCAACCCACATCACTAAGCCGACCGCAGTGGAGCGGCCCGGCAACAGCA 148
Db 129 TGTTACTTCAACCCACATCACTAAGCCGACCGCAGTGGAGCGGCCCGGCAACAGCA 188
Qy 149 GCAGTGTGGGAAAAAGGGGAGGAGCTGCGAGGCTGCGTCTGCGACCTGCTGG 208
Db 189 GCAGTGTGGGAAAAAGGGGAGGAGCTGCGAGGCTGCGTCTGCGACCTGCTGG 248
Qy 209 TGAAGCAGACGATGTCAGCGGCGCCCTGCTCTGCGGCGAGAGATCAACCGGACCA 268
Db 249 TGAAGCAGACGATGTCAGCGGCGCCCTGCTCTGCGGCGAGAGATCAACCGGACCA 308
Qy 269 AGGAGGAGGCGCTCGAGCTGATCAACGCTCATCTCAGAGATCAAGTTCGGGAGGAGG 328
Db 309 AGGAGGAGGCGCTCGAGCTGATCAACGCTCATCTCAGAGATCAAGTTCGGGAGGAGG 368
Qy 329 ACTTTGAGTCTCTGGCTCTCAGTTCAGGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 388
Db 369 ACTTTGAGTCTCTGGCTCTCAGTTCAGGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 428
Qy 389 TGGGTGCTTTCAGCAGAGGTCAGATGTCAGAGAGCCATTGAAGAGCGCTCTGCTGCTGC 448
Db 429 TGGGTGCTTTCAGCAGAGGTCAGATGTCAGAGAGCCATTGAAGAGCGCTCTGCTGCTGC 488
Qy 449 GGACGGGGGAGATGAGCGGGCCCGTGTTCACGGATTCGGGATTCACATCATCTCTCGCA 508
Db 489 GGACGGGGGAGATGAGCGGGCCCGTGTTCACGGATTCGGGATTCACATCATCTCTCGCA 548
Qy 509 CTGAGTCAGGCTGGGAGGCGCCAGCGCTGGCTCGGGCGAGGCGGGCGGTAGGCGGCG 568
Db 549 CTGAGTCAGGCTGGGAGGCGCCAGCGCTGGCTCGGGCGAGGCGGGCGGTAGGCGGCG 608
Qy 569 CAGTCTCCCGCTTGGCCGCGCAGCAGTGGCGGAAACCCCGCCACTCTCCCTGCGACCGTCACACA 628
Db 609 CAGTCTCCCGCTTGGCCGCGCAGCAGTGGCGGAAACCCCGCCACTCTCCCTGCGACCGTCACACA 668
Qy 629 GTATTATTGTTCCCAATGGCTGGGAGGGGCGCTTCCAGATGGGGGCGCTGGGGTTC 688
Db 669 GTATTATTGTTCCCAATGGCTGGGAGGGGCGCTTCCAGATGGGGGCGCTGGGGTTC 726
Qy 689 CCACATCCCTGCTCCATCCCAAGTGGGGCTCGGACCGCCAGATTCCTCCCTTAAGGAAATTG 748
Db 727 CCACATCCCTGCTCCATCCCAAGTGGGGCTCGGACCGCCAGATTCCTCCCTTAAGGAAATTG 785
Qy 749 ACTTCAGCAGGGGTGGGAGGCTCCCGAGACCGAGGCGAGTGTGGTGGAGGGGTGTTCAA 808
Db 786 ACTTCAGCAGGGGTGGGAGGCTCCCGAGACCGAGGCGAGTGTGGTGGAGGGGTGTTCAA 845
Qy 809 AGAAGAGGCTGTGTACAGAGAGCGCCCGCGTGTCCCCCGCCAGGTCGTGGAGCGAGCTCGA 868

Db 846 AGAAGAGGCTGTGTACAGAGAGCGCCCGTGTGCCCGCAGGTCTGCGAGGCGAGCTCGA 905
Qy 869 GGGCGGAATTTCTTAGTTAGGCGCAGCTCTCTCTGTTGAGTCGCGAAAGGTGAACACTCA 928
Db 906 GGGCGGAATTTCTTAGTTAGGCGCAGCTCTCTCTGTTGAGTCGCGAAAGGTGAACACTCA 965
Qy 929 TGGCG--CAGCCATGGGCGCTCTGAGCAACTGTGCGAGACCTTTTACCCCGCAATTAACCC 986
Db 966 TGGCGCCAGCATGGGCGCTCTGAGC-ACTGTGASAMCCTTTTACCCCGCAATTAACCA 1024
Qy 987 CAGAAC 992
Db 1025 GAACAC 1030

RESULT 2
AL526011/c
LOCUS
DEFINITION AL526011 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
CDNA clone CS0DC015YK14 3-PRIME, mRNA sequence.
ACCESSION AL526011
VERSION AL526011.2 GI:31063874
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 988)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT On Feb 13, 2001 this sequence version replaced gi:12789504.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9850.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DC015BF07NP1&cluster=9850.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DC015BF07NP1.

FEATURES

source
1..988
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC015YK14"
/tissue_type="NEUROBLASTOMA"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 89.7%; Score 891.4; DB 9; Length 988;
Best Local Similarity 98.2%; Pred. No. 8.2e-170;
Matches 907; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

Qy 29 CGGACGAGGAGAACTGCGCCGCTGGGAGAGCGCATGAGCGCAGCTCAGGCCGAG 88
Db 960 CGGACGAGGAGAACTGCGCCGCTGGGAGAGCGCATGAGCGCAGCTCAGGCCGAG 901
Qy 89 TGTTACTTCAACCCACATCACTAAGCCGACCGCAGTGGAGCGGCCCGGCAACAGCA 148
Db 900 TGTTACTTCAACCCACATCACTAAGCCGACCGCAGTGGAGCGGCCCGGCAACAGCA 841
Qy 149 GCAGTGTGGGAAAAAGGGGAGGAGCTGCGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 208

QY 613 CTGCGCCGTCACACAGTATTATTGTTCCACAAATGCTGGGAGGGGCGCTTCCAGAT 672
 Db 349 CTGCGCCGTCACACAGTATTATTGTTCCACAAATGCTGGGAGGGGCGCTTCCAGAT 290
 QY 673 TGGGGGCGCTGGGGTCCCGCACTCCCTGTCATCCCGGAGTTGGGGCTGCGAGCGGCAGATT 732
 Db 289 TGGGGGCGCTGGGGTCCCGCACTCCCTGTCATCCCGGAGTTGGGGCTGCGAGCGGCAGATT 230
 QY 733 CTCCCTTAAGAAATTGATTTGACAGAGGCTGGAGGCTCCAGACCGCGGAGGAGTGGT 792
 Db 229 CTCCCTTAAGAAATTGATTTGACAGAGGCTGGAGGCTCCAGACCGCGGAGGAGTGGT 170
 QY 793 GGGAGGGGCTGTTCCAAAGAGAGGCTGGTTCAGCAGAGCGCCCGGCTGTCGCCCGCAGGTG 852
 Db 169 GGGAGGGGCTGTTCCAAAGAGAGGCTGGTTCAGCAGAGCGCCCGGCTGTCGCCCGCAGGTG 110
 QY 853 CTGAGGAGCAGACTCGAGGGCGGAATTGTTCTAGTTAGGCGCACCTCTCTGTTGAGTGG 912
 Db 109 CTGAGGAGCAGACTCGAGGGCGGAATTGTTCTAGTTAGGCGCACCTCTCTGTTGAGTGG 50
 QY 913 CAAAGGTGAACATCATGCGC---AGCCATGGGCGCTCTGAGCAACT 957
 Db 49 CAAAGGTGAACATCATGCGC---AGCCATGGGCGCTCTGAGCAACT 2

RESULT 4
 BX436611 1076 bp mRNA linear EST 15-MAY-2003
 LOCUS BX436611 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP002YN11
 DEFINITION 5-PRIME, mRNA sequence.

ACCESSION BX436611
 VERSION BX436611.1 GI:30785532
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1076)
 Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 9850.f For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0CAP002CG06QPI&cluster=9850.f. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0CAP002CG06QPI.

FEATURES
 source
 1..1076
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0CAP002YN11"
 /tissue_type="THYMUS"
 /clone_lib="Homo sapiens THYMUS"
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo(dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

ORIGIN

Query Match 86.4%; Score 859; DB 13; Length 1076;
 Best Local Similarity 94.0%; Pred. No. 3e-163;
 Matches 911; Conservative 21; Mismatches 31; Indels 6; Gaps 4;
 QY 31 GACGAGGAGAGTGTCCCGCGCTGGGAGAGCGCATGAGCCGCGGCGGAGTG 90

Db 83 GTCCGGAGAGAGTGTCCCGCGCTGGGAGAGAGCGCATGAGCCGAGCTCAGGCGAGTG 142
 QY 91 TACTACTTCAACCACTACATCACTAACGCCAGCGCAGTGGGAGCGGCCAGCGCAACAGCAGC 150
 Db 143 TACTACTTCAACCACTACATCACTAACGCCAGCGCAGTGGGAGCGGCCAGCGCAACAGCAGC 202
 QY 151 AGTGGTGGCAAAACGGGAGGGGAGCTGCCAGGGTCCGCTGCTGCGACCTGCTGGTG 210
 Db 203 AGTGGTGGCAAAACGGGAGGGGAGCTGCCAGGGTCCGCTGCTGCGACCTGCTGGTG 262
 QY 211 AAGCAGACCCAGTCAACCGCGGCTCTGCTGCGGGGAGGAGATCAACCGGACCAAG 270
 Db 263 AAGCAGACCCAGTCAACCGCGGCTCTGCTGCGGGGAGGAGATCAACCGGACCAAG 322
 QY 271 GAGGAGCCCTGGAGCTGATCAACCGGCTAATCAGAGAGATCAAGTCCGGAGAGGAGAC 330
 Db 323 GAGGAGCCCTGGAGCTGATCAACCGGCTAATCAGAGAGATCAAGTCCGGAGAGGAGAC 382
 QY 331 TTTGAGTCTCTGGCTCAACGCTTTCAGGAGCTGAGCTCAGCCAAAGGCCAGGGGAGACCTG 390
 Db 383 TTTGAGTCTCTGGCTCAACGCTTTCAGGAGCTGAGCTCAGCCAAAGGCCAGGGGAGACCTG 442
 QY 391 GGTGCTTTGAGCAGAGGTGATGAGAGGCAATTTGAAGACGCTCTGCTGCTGGTG 450
 Db 443 GGTGCTTTGAGCAGAGGTGATGAGAGGCAATTTGAAGACGCTCTGCTGCTGGTG 502
 QY 451 ACGGGGAGAGTACGGGGCGGCTGTTACGAGATTCGGGATCCACATCATCTCTCGCACT 510
 Db 503 ACGGGGAGAGTACGGGGCGGCTGTTACGAGATTCGGGATCCACATCATCTCTCGCACT 562
 QY 511 GAGTGAGGGTGGGAGCGCCAGGCTCGGCTCGGGCAGGGCGGCTAGGGCGGCCA 570
 Db 563 GAGTGAGGGTGGGAGCGCCAGGCTCGGCTCGGGCAGGGCGGCTAGGGCGGCCA 622
 QY 571 GCTCCCCCTTCCCGCCAGCCAGTGGCGGAGACCCCGGAGATTCCTTCAAGGAAATTCAC 630
 Db 623 GCTCCCCCTTCCCGCCAGCCAGTGGCGGAGACCCCGGAGATTCCTTCAAGGAAATTCAC 682
 QY 631 ATTTATTGTTCCCAATGGCTGGGAGGGGGGCGCTTCCAGATTTGGGGGCGCTGGGTCCC 690
 Db 683 ATTTATTGTTCCCAATGGCTGGGAGGGGGCGCTTCCAGATTTGGGGGCGCTGGGTCCC 742
 QY 691 CACTCCCTGTCCTATCCCGAGTGGGGCTGGACCGCCAGATTCCTTCAAGGAAATTCAC 750
 Db 742 CACTCCCTGTCCTATCCCGAGTGGGGCTGGACCGCCAGATTCCTTCAAGGAAATTCAC 801
 QY 751 TTCAGCGGGTGGGAGGCTCCAGACCCAGGGCAGTGTGGT-GGGAGGGGTGTTCGAAA 809
 Db 802 TTCAGCGGGTGGGAGGCTCCAGACCCAGGGCAGTGTGGTGGGGAGGGGTGTTCGAAA 861
 QY 810 GAGAGGCTGGTTCAGCAGAGCGCCCGCTGTCGCCAGGTGTGGAGGCGAGACTCCAG 869
 Db 862 GAGAGGCTGGTTCAGCAGAGCGCCCGCTGTCGCCAGGTGTGGAGGCGAGACTCCAG 921
 QY 870 GGGCGAATTTGTTCTAGTTAGGCGCACGCTCTCTGTTTTCAGTTCGAAAAGTCAACTCAT 929
 Db 922 GGGCGAATTTTCTAGTTAGGCGCACGCTCTCTGTTTTCAGTTCGAAAAGTCAACTCAT 981
 QY 930 GCGGCG---AGCCATGGGCGCTCTGAGCAACTGTGAGACCTTTTCA-CCCCCAATTAAC 985
 Db 982 KCGGCCCAACATAGSSGCTCGTGAGCAAAVKTGTCAACCCGKKKATMCCCCCAAKTMAA 1041
 QY 986 CCAGAACCA 994
 Db 1042 CCAAAACMA 1050

RESULT 5

BI829812

LOCUS

DEFINITION

BI829812 923 bp mRNA linear EST 04-OCT-2001
 603079887F1 NIH_MGC_1:9 Homo sapiens cDNA clone IMAGE:5171540 5',
 mRNA sequence.


```

/dev stage="adult"
/clone.lib="Homo sapiens ADULT BRAIN"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN
Query Match      82.2%; Score 817.2; DB 13; Length 1201;
Best Local Similarity 99.3%; Pred. No. 9e-155;
Matches 842; Conservative 0; Mismatches 3; Indels 3; Gaps 2;

QY 150 CAGTGGTGGCAAAACGGGCGAGGCTGCGAGGGTCCGCTGCTGCACCTGCTGCT 209
DB 60 CCGGATGTCAAAACGGGCGAGGAGCTGCGAGGGTCCGCTGCTGCACCTGCTGCT 119

QY 210 GAAGCACACCCAGTCAAGCGCGCCCTGCTGCTGGCGGCGAGGAAGATCACCGGACCAA 269
DB 120 GAAGCACACCCAGTCAAGCGCGCCCTGCTGCTGGCGGCGAGGAAGATCACCGGACCAA 179

QY 270 GAGAGAGCCCTGGAGCTCATCAAGCTTACATCCAGAGATCAAGTCCGGAGAGGAGA 329
DB 180 GAGAGAGCCCTGGAGCTCATCAAGCTTACATCCAGAGATCAAGTCCGGAGAGGAGA 239

QY 330 CTTTCAAGTCTTGGGCTCAAGTTCAGGCTGCTGAGCTCAGCTCAGCCNAGCCAGGAGACCT 389
DB 240 CTTTCAAGTCTTGGGCTCAAGTTCAGGCTGCTGAGCTCAGCTCAGCCNAGCCAGGAGACCT 299

QY 390 GGGTGCCTTCAGCAGAGGTCAATGACAGAGCCATTTGAAGAGCGCTGCTGTTGGCTGGC 449
DB 300 GGGTGCCTTCAGCAGAGGTCAATGACAGAGCCATTTGAAGAGCGCTGCTGTTGGCTGGC 359

QY 450 GACGGGGAGATGACGGGCGCGTGTTCAGGATTCGGCATCCGATCCATCATCTCTCGCAC 509
DB 360 GACGGGGAGATGACGGGCGCGTGTTCAGGATTCGGCATCCGATCCATCATCTCTCGCAC 419

QY 510 TCAGTTCAGGCTGGGAGCCAGGCTGCTGCGGCGAGGAGGAGGCGCTAGGCGCGCC 569
DB 420 TGAAGTGGGGTGGGAGCCAGGCTGCTGCGGCGAGGAGGAGGCGCTAGGCGCGCC 479

QY 570 AGCTCCCTTCGCGCGAGGAGTGGCGAGAGCCCGGCTGCTGCTGCGCGAGGAGTGA 629
DB 480 AGCTCCCTTCGCGCGAGGAGTGGCGAGAGCCCGGCTGCTGCTGCGCGAGGAGTGA 539

QY 630 TATTTATTTCTCCCAATGGCTGGGAGGGGCGCTTCCAGATTGGGGGCGCTGGGGTCC 689
DB 540 TATTTATTTCTCCCAATGGCTGGGAGGGGCGCTTCCAGATTGGGGGCGCTGGGGTCC 599

QY 690 CCACTCCCTTCATCCCGAGTGGGGCTCCGACCGCGAGATTCCTCCCTTAAGGAATTGA 749
DB 600 CCACTCCCTTCATCCCGAGTGGGGCTCCGACCGCGAGATTCCTCCCTTAAGGAATTGA 659

QY 750 CTTTACAGGGGTGGAGGCTCCAGACCCAGGCGAGTGTGGTGGAGGGGTGTTCCAA 809
DB 660 CTTTACAGGGGTGGAGGCTCCAGACCCAGGCGAGTGTGGTGGAGGGGTGTTCCAA 719

QY 810 GAGAGGCTGTGTACAGAGCGCGCCGCTGCTCCCGAGTGTGTGGAGGAGACTCGAG 869
DB 720 GAGAGGCTGTGTACAGAGCGCGCCGCTGCTCCCGAGTGTGTGGAGGAGACTCGAG 779

QY 870 GCGCGAATTTTCTAGTTAGGCGACGCTCCTCTGTTTCACTGCGCAAGGTTGAACTCAT 929
DB 780 GCGCGAATTTTCTAGTTAGGCGACGCTCCTCTGTTTCACTGCGCAAGGTTGAACTCAT 839

QY 930 GCGG--CAGCCATGGGCGCTCTGAGCACTGTGAG--ACCTTTTCACCCCCAATTAAAC 986
DB 840 GCGGCGGAGCATGGGCGCTCTGAGCACTGTGAGCACTGTGAGCACTGTGAGCACT 899

QY 987 CAGAACCA 994
DB 900 CAGAACCA 907

```

RESULT 7

AL529668 1097 bp mRNA linear EST 23-MAY-2003
 AL529668 Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED Homo sapiens
 CDNA clone CS0DD005YA23 3-PRIME, mRNA sequence.

ACCESSION
 VERSION
 AL529668
 EST.

SOURCE
 Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1097)

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

On Feb 13, 2001 this sequence version replaced gi:12793161.

Contact: Genoscope

Genoscope - Centre National de Sequencage

Bp 191 91006 BVRV cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

was normalized. Library was constructed by life technologies, a

division of invitrogen. This sequence belongs to sequence cluster

9850.f For more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0DD005AA12NP1&cluster=9850.f. Contact :

Peng Liang Email : fliang@lifetech.com URL :

http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Paradise Avenue Genoscope sequence ID : CS0DD005AA12NP1.

Location/Qualifiers

1..1097

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CS0DD005YA23"

/tissue type="NEUROBLASTOMA COT 50-NORMALIZED"

/clone lib="Homo sapiens NEUROBLASTOMA COT 50-NORMALIZED"

/note="1st strand cDNA was primed with a NotI-oligo(dT)

primer. Five prime end enriched, double-strand cDNA was

digested with Not I and cloned into the Not I and EcoRV

sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 81.1%; Score 806.2; DB 9; Length 1097;
 Best Local Similarity 91.4%; Pred. No. 1.5e-152;
 Matches 884; Conservative 15; Mismatches 39; Indels 29; Gaps 4;

QY 11 CACCTCAGGGAAGATGGCGGACGAGGAGAGTGGCGGCGGCTGGGAGAGCGCATGA 70
 DB 987 CCGGSRGTGGGAGGNTGGCGGACGAGAGAGTGTGGCGGCTGTGTTGAAATGCGCTGA 928
 QY 71 GCGCAGCTCAGCGCGAGTGTACTTCAACACATCACTAACCGCAGCAGTGGGAGC 130
 DB 927 GCGCGGCTCAGCGCGAGTGNACATTAACACACACACACACACACACACAC 869
 QY 131 GCGCCAGCGGACACAGCAGCAGTGGTGGCAAAACCGGCGAGGAGGAGCTGCCAGGTTCC 190
 DB 868 GCGCCAGCGGACACAGCAGCAGTGGTGGCAAAACCGGCGAGGAGGAGCTGCCAGGTTCC 810
 QY 191 GCTGCTGACCTGCTGCTGTAAGCAGCAGCAGTTCAGCGGCGGCTGCTGCGGCGAG 250
 DB 809 GCTGCTGACCTGCTGCTGTAAGCAGCAGCAGTTCAGCGGCGGCTGCTGCGGCGAG 750
 QY 251 AGAAGATCACCCTGACCAAGGAGGAGCGCTGCGAGTGTATCAACGGGTACATCCAGAAGA 310
 DB 749 AGAAGATCACCCTGACCAAGGAGGAGCGCTGCGAGTGTATCAACGGGTACATCCAGAAGA 690
 QY 311 TCAAGTGGGAGGAGGAGCTTTGAGTCTTGGGCTTCAAGTTCAGGAGTTCAGGAGTTCAG 370
 DB 689 TCAAGTGGGAGGAGGAGCTTTGAGTCTTGGGCTTCAAGTTCAGGAGTTCAGGAGTTCAG 630
 QY 371 CCAAGGCGGAGGAGGAGCTGCTGCTTTCAGCAG-----A 405

```
Db 629 CCAAGCCAGGGGAGACCTGGGTCCTTCAGCAGAGGTGCGAAGGATGGGCTCACCA 570
QY 406 GGTTCAGATGAGAGCCATTGAGAGCGCTTTCGCTTTCGCTGCGGAGCGGGGAGATGAGC 465
Db 569 GGTTCAGATGAGAGCCATTGAGAGCCCTTGGTTTCGCTTTCGCTGCGGAGCGGGGAGTGGG 510
QY 466 GGGCCCGTGTTCAGAGTTCGGGATCCACATCATCTCCGCACTGAGTGAGGAGTGGGGA 525
Db 509 GGGCCCGGKGTTCAGAGTTCGGGATCCACATCATCTCCGCACTGAGTGAGGAGTGGGGA 450
QY 526 GCCCAGCGCTGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 585
Db 449 GCCCAGCGCTGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 390
QY 586 CCAGCAGTGGGCGAGCCCGCCCTCCCTCCGACCGTCCACAGTATTTATTTGTTCCAC 645
Db 389 CCAGCAGTGGGCGAGCCCGCCCTCCCTCCGACCGTCCACAGTATTTATTTGTTCCAC 330
QY 646 AATGGCTGGGAGGGGCGCTTCCAGATTTGGGAGCGGCTTCCGAGTGGGAGCGGAGCGGAG 705
Db 329 AATGGCTGGGAGGGGCGCTTCCAGATTTGGGAGCGGCTTCCGAGTGGGAGCGGAGCGGAG 270
QY 706 CCAGTGGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 765
Db 269 CCAGTGGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 210
QY 766 AGGCTCCAGAGCCCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 825
Db 209 AGGCTCCAGAGCCCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 150
QY 826 CAGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 885
Db 149 CAGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 90
QY 886 GTTAGCCAGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 943
Db 89 GTTAGCCAGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 30
QY 944 CCTCTG 950
Db 29 CCTCTG 23
```

```
RESULT 8
AL533491
LOCUS
DEFINITION
AL533491 Homo sapiens ADULT BRAIN Homo sapiens cDNA clone
CS0DN004YN04 5-PRIME, mRNA sequence.
ACCESSION
AL533491
VERSION
AL533491.2 GI:31260572
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 868)
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
On Feb 13, 2001 this sequence version replaced gi:12796984.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 9850.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DN004DG020P1&cluster=9850.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ InvitroGen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DN004DG020P1.
```

```
FEATURES
source
1..868
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DN004YN04"
/tissue_type="ADULT BRAIN"
/dev stage="adult"
/clone_lib="Homo sapiens ADULT BRAIN"
/note="Organ: Brain; vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
ORIGIN
Query Match 81.0%; Score 805.4; DB 9; Length 868;
Best Local Similarity 99.1%; Pred. No. 1.9e-152;
Matches 806; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 13 CCTCGAGGAGAGTGGCGAGGAGAGAGTGGCGGCGGCTGGGAGAGCGATGAGC 72
Db 56 CCGGATGGAGATGGCGGAGCGAGGAGAGTGGCGGCGGCTGGGAGAGCGATGAGC 115
QY 73 CGCAGCTCAGGCGGAGTGTACTTCAACCAATCACTAACGCCAGCCAGTGGGAGCGG 132
Db 116 CGCAGCTCAGGCGGAGTGTACTTCAACCAATCACTAACGCCAGCCAGTGGGAGCGG 175
QY 133 CCAGCGGCAACAGCAGCAGTGTGTGCAAAAACCGGCGAGGAGCGTGGCAGGCTCCGC 192
Db 176 CCAGCGGCAACAGCAGCAGTGTGTGCAAAAACCGGCGAGGAGCGTGGCAGGCTCCGC 235
QY 193 TGCTCGCAGCTGCTGTGAGCAGCAGCAGTGTGTGAGCAGCGGCGGCTGCTGCTGCGCAGGAG 252
Db 236 TGCTCGCAGCTGCTGTGAGCAGCAGCAGTGTGTGAGCAGCGGCGGCTGCTGCTGCGCAGGAG 295
QY 253 AAGATCACCGGACCAAGAGGAGGCGCTCGAGCTGATCAACGGCTACATCCAGAGATC 312
Db 296 AAGATCACCGGACCAAGAGGAGGCGCTCGAGCTGATCAACGGCTACATCCAGAGATC 355
QY 313 AAGTCGGAGAGAGGAGCTTTGAGTCTTGGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 372
Db 356 AAGTCGGAGAGAGGAGCTTTGAGTCTTGGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 415
QY 373 AAGTCGGAGAGAGGAGCTTTGAGTCTTGGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 432
Db 416 AAGTCGGAGAGAGGAGCTTTGAGTCTTGGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT 475
QY 433 GCTCGCTTTGCGCTGCGGAGAGATGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 492
Db 476 GCTCGCTTTGCGCTGCGGAGAGATGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 535
QY 493 CACATCATCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552
Db 536 CACATCATCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 595
QY 553 GGGCGCTAGGCGGCGGAGCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 612
Db 596 GGGCGCTAGGCGGCGGAGCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 655
QY 613 CTGCCACCGTCAACAGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 672
Db 656 CTGCCACCGTCAACAGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 715
QY 673 TGGGGGCGCTTGGGCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 732
Db 716 TGGGGGCGCTTGGGCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 775
QY 733 CTCCCTTAAAGGAATTTAGCTTTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 792
Db 776 CTCCCTTAAAGGAATTTAGCTTTCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 835
QY 793 GGGAGGGGTGTTCCAAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 825
```



```
/tissue type="T CELLS (JURKAT CELL LINE)"
/cell_line="JURKAT CELL LINE"
/clone_lib="Homo sapiens T CELLS (JURKAT CELL LINE),"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
```

ORIGIN

```
Query Match 79.8%; Score 792.8; DB 13; Length 1201;
Best Local Similarity 97.0%; Pred. No. 7.7e-150;
Matches 844; Conservative 6; Mismatches 13; Indels 7; Gaps 4;

QY 13 CCTCGAGGAGATGGCGGAGGAGAGCTGGCGCGCGCTGGGAGCGCATGAGC 72
DB 43 CCGGGATGGAGATGGCGGAGGAGCTGGCGCGCGCTGGGAGCGCATGAGC 102
QY 73 CGCAGCTCAGGCGGAGTGTACTTCAACACATCACTAACCGCAGCCAGTGGGAGCGG 132
DB 103 CGCAGCTCAGGCGGAGTGTACTTCAACACATCACTAACCGCAGCCAGTGGGAGCGG 162
QY 133 CCGAGGCGCAACAGCAGCGTGTGGGAAAAACGGGCGAGGGGAGCCCTGCAGAGTCCGC 192
DB 163 CCGAGGCGCAACAGCAGCGTGTGGGAAAAACGGGCGAGGGGAGCCCTGCAGAGTCCGC 222
QY 193 TGCTCGCAGCTGCTGGTGAAGCAGCAGCCAGTCAO3GGCGCCCTGCTCTGCGCGGAGG 252
DB 223 TGCTCGCAGCTGCTGGTGAAGCAGCAGCCAGTCAO3GGCGCCCTGCTCTGCGCGGAGG 282
QY 253 AAGATCACCCGAGCAAGGAGGAGCCCTGGAGCTGATCAACGGCTACATCCAGAGATC 312
DB 283 AAGATCACCCGAGCAAGGAGGAGCCCTGGAGCTGATCAACGGCTACATCCAGAGATC 342
QY 313 AAGTCGAGGAGGAGGACTTTGAGTCTCTGGCTCAGATTCAGGAGTGGAGTTCAGCTCAGCC 372
DB 343 AAGTCGAGGAGGAGGACTTTGAGTCTCTGGCTCAGATTCAGGAGTGGAGTTCAGCTCAGCC 402
QY 373 AAGGCCAGGAGGAGACTGGGTGCTTCAGCAGAGTTCAGTTCAGAGCCATTTGAAGAC 432
DB 403 AAGGCCAGGAGGAGACTGGGTGCTTCAGCAGAGTTCAGTTCAGAGCCATTTGAAGAC 462
QY 433 GCCTCGTTTGGCTGCGGAGCGGAGATGAGCGGCGCCCGTGTTCACG3ATTTCGGCATC 492
DB 463 GCCTCGTTTGGCTGCGGAGCGGAGATGAGCGGCGCCCGTGTTCACG3ATTTCGGCATC 522
QY 493 CACATCATCTCCGACTGAGTGAAGGTGGGAGCCAGAGGCTTGGCTTCGGGCGAGGCA 552
DB 523 CACATCATCTCCGACTGAGTGAAGGTGGGAGCCAGAGGCTTGGCTTCGGGCGAGGCA 582
QY 553 GGGCGGCTAGGCGCGGAGCTCCCTTGGCGCGCAGTGGCGGAAACCCCACTCC 612
DB 583 GGGCGGCTAGGCGCGGAGCTCCCTTGGCGCGCAGTGGCGGAAACCCCACTCC 642
QY 613 CTGCACCGCTCACACAGTATTTATTTGTTCCCAATGGCTG3GAGGGGCGCTTCCAGAT 672
DB 643 CTGCACCGCTCACACAGTATTTATTTGTTCCCAATGGCTG3GAGGGGCGCTTCCAGAT 702
QY 673 TGGGGGCGCTGGGGTTCCTCATCTCCCTGTCATCCAGTGGGCGTGGACCGCCAGATT 732
DB 703 TGGGGGCGCTGGGGTTCCTCATCTCCCTGTCATCCAGTGGGCGTGGACCGCCAGATT 762
QY 733 CTCCCTTAAGGAATTGACTTCAGCAGAGGTGGAGGCTCCACAGCCAGGGCAGTGTGT 792
DB 763 CTCCCTTAAGGAATTGACTTCAGCAGAGGTGGAGGCTCCACAGCCAGGGCAGTGTGT 822
QY 793 GGGAGGGGTGTTCAAAAGAGAGGCTGTGTTCAGCAGAGCGCGCCCGTGTCCCCCAGGTG 852
DB 823 GGGAGGGGTGTTCAAAAGAGAGGCTGTGTTCAGCAGAGCGCGCCCGTGTCCCCCAGGTG 876
QY 853 CTGAGGCGAGCTCGAGGGCGCAATTGTTT 882
DB 877 CTGGA-GSAGACTCGAGGGCGCAATTGTTT 905
```

RESULT 11

```
BUS57229
LOCUS BUS57229
DEFINITION AGNCOURT_102533322 NIH_MGC_109 Homo sapiens cDNA clone
IMAGE:6585062 5', mRNA sequence.
ACCESSION BUS57229
VERSION BUS57229.1 GI:22907525
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
MAMMALIA: Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
MAMMALIA: Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 864)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed By: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2796 row: g column: 14
High quality sequence stop: 728.
FEATURES
location/Qualifiers
1..864
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6585062"
/tissue_type="teratocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_109"
/notes="Organ: ovary; Vector: pOTB7; Site 1: EcoRI; Site 2:
XhoI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
```

ORIGIN

```
Query Match 78.9%; Score 784.2; DB 13; Length 864;
Best Local Similarity 97.0%; Pred. No. 3.7e-148;
Matches 819; Conservative 0; Mismatches 21; Indels 4; Gaps 2;

QY 8 CAGCACCTCGAGGGAAGATGGCGGAGGAGAGCTGGCGCGCGCTGGGAGAGCGCA 67
DB 5 CTGGCGGAGGAGGGAAGATGGCGGAGGAGAGCTGGCGCGCGCTGGGAGAGCGCA 64
QY 68 TGAGCGCGCAGCTCAGCGCGAGTGTACTTCAACACATCACTAACCGCAGCGAGTGG 127
DB 65 TGAGCGCGCAGCTCAGCGCGAGTGTACTTCAACACATCACTAACCGCAGCGAGTGG 124
QY 128 AGCGGCCCGAGCGCAACAGCAGCAGTGTGGCAAAACGGGCGAGGGGAGCCTGCCAGG 187
DB 125 AGCGGCCCGAGCGCAACAGCAGCAGTGTGGCAAAACGGGCGAGGGGAGCCTGCCAGG 184
QY 188 TCCGCTGTCTCGACCTGTGTGAGCAGCAGCAGTGTGGCGCGCGCTGTCTGTGGCGC 247
DB 185 TCCGCTGTCTCGACCTGTGTGAGCAGCAGCAGTGTGGCGCGCGCTGTCTGTGGCGC 244
QY 248 AGGAGAAGATCAACCGGACCAAGGAGGAGCGCCCTGGAGCTGATCAACGCTACATCCAG 307
DB 245 AGGAGAAGATCAACCGGACCAAGGAGGAGCGCCCTGGAGCTGATCAACGCTACATCCAG 304
QY 308 AGATCAAGTCG3GAGAGGAGGAGCTTTGAGTCTCTGTCTGTCTGTCTGTCTGTCTGT 367
```

```

Db      305 AGATCAAGTGGGAGAGAGAGACTTTGAGTCTCTGGCCCTCAAGTTACGACATGACGCT 364
QY      368 CAGCAAGGGCCAGGGAGAGACTTGGTGCTCTTACAGAGAGGTGAGATGCGAAGCAATTG 427
Db      365 CAGCAAGGGCCAGGGAGAGACTTGGTGCTCTTACAGAGAGGTGAGATGCGAAGCAATTG 424
QY      428 AAGACGCTCTGTTGGCTGGAGCGGGAGAGATGAGCGGCGCGTGTTCAGGATTCG 487
Db      425 AAGACGCTCTGTTGGCTGGAGCGGGAGAGATGAGCGGCGCGTGTTCAGGATTCG 484
QY      488 GCATCCACATCATCTCCGCACTGAGTGAAGGTGGGAGCGCCAGGCTGGCTCGGGCA 547
Db      485 GCATCCACATCATCTCCGCACTGAGTGAAGGTGGGAGCGCCAGGCTGGCTCGGGCA 544
QY      548 GGGCAGGGCGCTAGGGCGGCGAGCTCCCTCTGCGCGCGAGCGAGTGGCGAAGCGGCC 607
Db      545 GGGCAGGGCGGCTAGGGCGGCGAGCTCCCTCTGCGCGCGAGCGAGTGGCGAAGCGGCC 604
QY      608 ACTCCCTGCCACCGTCAACAGATATTATTGTTCCCAATGGCTGGGAGGGGGCCCTTC 667
Db      605 ACTCCCTGCCACCGTCAACAGATATTATTGTTCCCAATGGCTGGGAGGGGGCCCTTC 664
QY      668 CAGATTGGGGCGGCTCGGGTCCCACTCCCTGTCCATCCCGAGTTGGGGCTGGACCGCC 727
Db      665 CAGATTGGGGCGGCTCGGGTCCCACTCCCTGTCCATCCCGAGTTGGGGCTGGACCGCC 724
QY      728 AGATTCTCCCTTAAGGAAATGACTTCAGCAGGGGTGGAGGCTCCAGACCGAGGAGT 787
Db      725 AGATTCTCCCTTAAGGAAATGACTTCAGCAGGGGTGGAGGCTCCAGACCGAGGAGT 784
QY      788 GTGGT--GGGAGGGGTGTTCCAAAGAGAA--GGGCTGTGAGCAGAGCGCGCGCTGTCC 843
Db      785 GTGGTGGGAGGGGTGTTCCAAAGAGAAAGGCTGTGATCAGCAGAGCGCGCGCTGT 844
QY      844 CCCC 847
Db      845 CCCC 848

```

RESULT 12

BQ678260
LOCUS BQ678260 928 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_841988 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6274055
5', mRNA sequence.

ACCESSION BQ678260
VERSION BQ678260.1 GI:21790939

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 928)

REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cga@nci.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.llnl.gov>

Plate: LLM2454 row: p column: 24

High quality sequence stop: 685.

FEATURES

source

/tissue type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_112"

/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GCCACGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN

```

Query Match      78.6%; Score 781; DB 13; Length 928;
Best Local Similarity 96.5%; Pred. No. 1.7e-147;
Matches 831; Conservative 0; Mismatches 25; Indels 5; Gaps 3;
QY 28 GCGCAGGAGAGAGCTGCCCGCGGTGGGAGAGCGCATGAGCGCGAGCTCAGGCCGA 87
Db 1 GCGCAGGAGAGAGCTGCCCGCGGTGGGAGAGCGCATGAGCGCGAGCTCAGGCCGA 60
QY 88 GTGTACTACTTTAAACACATCACTAAACAGCAGTGGGAGCGGCCAGCGGCAACAGC 147
Db 61 GTGTACTACTTTAAACACATCACTAAACAGCAGTGGGAGCGGCCAGCGGCAACAGC 120
QY 148 AGCAGTGTGTGCAAAAGCGGCGAGGCGAGCTCCAGGGTCCGCTGCTCCACCTGCTG 207
Db 121 AGCAGTGTGTGCAAAAGCGGCGAGGCGAGCTCCAGGGTCCGCTGCTCCACCTGCTG 180
QY 208 GTGAAGCAGCAGCCAGTCACTCCGCGGCTCGCTCTGGGCGCAGGAGAGATCACCGGACC 267
Db 181 GTGAAGCAGCAGCCAGTCACTCCGCGGCTCGCTCTGGGCGCAGGAGAGATCACCGGACC 240
QY 268 AAGSAGAGAGCGCTGGAGCTGATCAAGCGCTACATCCAGAGATCAAGTCGGGAGAGAG 327
Db 241 AAGSAGAGAGCGCTGGAGCTGATCAAGCGCTACATCCAGAGATCAAGTCGGGAGAGAG 300
QY 328 GACTTTCAGTCTCTGGCTTCACAGTTCAGCGACTGACGCTCAGCAAGGCGCAGGGAGAG 387
Db 301 GACTTTCAGTCTCTGGCTTCACAGTTCAGCGACTGACGCTCAGCAAGGCGCAGGGAGAG 360
QY 388 CTGGTCTCTTCAAGCAGAGTCAAGATCAGAAAGCCATTGAAAGAGCGCTCGTTTGGCTG 447
Db 361 CTGGTCTCTTCAAGCAGAGTCAAGATCAGAAAGCCATTGAAAGAGCGCTCGTTTGGCTG 420
QY 448 CCGACGGGGGAGATGAGCGGGCGCGTTCACGAGATTCGGGCGTCCAGATCATCATCTCCG 507
Db 421 CCGACGGGGGAGATGAGCGGGCGCGTTCACGAGATTCGGGCGTCCAGATCATCATCTCCG 480
QY 508 ACTGAGTGAAGGTGGGAGAGCGCTCGGCTCGGGCGCAGGGCAGGCGGCTAGGGCGG 567
Db 481 ACTGAGTGAAGGTGGGAGAGAGCGCTCGGCTCGGGCGCAGGGCAGGCGGCTAGGGCGG 540
QY 568 CAGCTCCCGCTGCGCGCCAGCGAGTGGCGGAGAAACCCCGCAGCTCCCTGCGACCGTCAAC 627
Db 541 CAGCTCCCGCTGCGCGCCAGCGAGTGGCGGAGAAACCCCGCAGCTCCCTGCGACCGTCAAC 600
QY 628 AGTATTTATTTGTTCCCAATGCTGGAGGGGCGCTCCAGATTCGGGCGCGCTGGGCT 687
Db 601 AGTATTTATTTGTTCCCAATGCTGGAGGGGCGCTCCAGATTCGGGCGCGCTGGGCT 660
QY 688 CCCCACCTCCCTGTCATCCCGCAGTTGGGCTGGAGCCCGCAGAGTTCTCCCTTAAGGAATT 747
Db 661 CCCCACCTCCCTGTCATCCCGCAGTTGGGCTGGAGCCCGCAGAGTTCTCCCTTAAGGAATT 720
QY 748 GACTTCAGAGGGGTGGAGGCTCCCGAGCCAGGGGAGTGTGGT--GGAGGGGTGTTC 805
Db 721 GACTTCAGAGGGGTGGAGGCTCCCGAGCCAGGGGAGTGTGGT--GGAGGGGTGTTC 780
QY 806 CAAAGAGAGAGCGCTGGT--CAGCAGAGCGCGCGCGCTGTCGCCCGAGGTGTGGAGGCA-G 862
Db 781 CAAAGAGAGAGCGCTGGT--CAGCAGAGCGCGCGCGCTGTCGCCCGAGGTGTGGAGGCA 840

```



```

QY 863 ACTCAGGGCGGAAATTTTC 863
Db 841 ACTCAGGGCGGAAATTTTC 861

RESULT 13
LOCUS BG744294
DEFINITION 602723212F1 NIH_MGC_106 Homo sapiens cDNA clone IMAGE:4849587 5',
mRNA sequence.
ACCESSION BG744294
VERSION B3744294.1 GI:14054947
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 833)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM1689 row: h column: 04
High quality sequence stop: 752.
Location/Qualifiers
1..833
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4849587"
/tissue_type="natural killer cells, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 106"
/note="Organ: blood; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."

ORIGIN
Query Match 78.1%; Score 776.8; DB 12; Length 833;
Best Local Similarity 98.1%; Pred. No. 1.1e-146;
Matches 806; Conservative 0; Mismatches 14; Indels 2; Gaps 2;

QY 8 CAGCACTTCGAGGAGATGGCGGACGAGAGAGCTGCCGCCGCTGGGAGACGCA 67
Db 2 TCGCGGAGAGGAGATGGCGGACGAGAGAGCTGCCGCCGCTGGGAGACGCA 61

QY 68 TGAGCGGAGCTCAGCGCCAGGTGTACTTCAACCAATCACTAACGCCAGCCAGTGGG 127
Db 62 TGAGCGGAGCTCAGCGCCAGGTGTACTTCAACCAATCACTAACGCCAGCCAGTGGG 121

QY 128 AGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAACCGGAGGGAGGCTGCCAGGG 187
Db 122 AGCGGCCAGCGGCAACAGCAGCAGTGTGGCAAAACCGGAGGGAGGCTGCCAGGG 181

QY 188 TCAGTGTCTCGCACTGTGTGAGCAGCAGCAGTCAAGCGGGCCCTCTGCTGGCGGC 247
Db 192 TCAGTGTCTCGCACTGTGTGAGCAGCAGCAGTCAAGCGGGCCCTCTGCTGGCGGC 241

QY 248 AGGAGAGATCACCGGACCAAGGAGGAGGCGCTCGAGGTGATCAAGCGGTACATCCAGA 301
Db 841 ACTCAGGGCGGAAATTTTC 861

```

```

Db 242 AGGAGAGATCCCCCGGACCAAGGAGGAGGCGCTGGAGTGATCAACGGCTACATCCAGA 301
QY 308 AGATCAAGTCCGGAGGAGGAGCTTTGAGTCTCTGGCCCTCAGAGTTACGAGTGCAGCT 367
Db 302 AGATCAAGTCCGGAGGAGGAGCTTTGAGTCTCTGGCCCTCAGAGTTACGAGTGCAGCT 361
QY 368 CAGCAAGCGCCAGGGGAGACCTGGGTGCTTCAGCAGAGGTGAGATGACAGAGCCATTTG 427
Db 362 CAGCAAGCGCCAGGGGAGACCTGGGTGCTTCAGCAGAGGTGAGATGACAGAGCCATTTG 421
QY 428 AAGACGCTCTGTTGGCTGCGGACGGGGAGATGAGCGGGCCCTGTTCAGGATTCG 487
Db 422 AAGACGCTCTGTTGGCTGCGGACGGGGAGATGAGCGGGCCCTGTTCAGGATTCG 481
QY 488 GCATCCCATCATCTCCGCACTGAGTGAGGGTGGGAGGCCAGCCCTGGCCCTCGGGCA 547
Db 482 GCATCCCATCATCTCCGCACTGAGTGAGGGTGGGAGGCCAGCCCTGGCCCTCGGGCA 541
QY 548 GGGCAGGGCGGCTAGGGCGGCGCAGTCCGCCCTTGGCCCGCCAGCCAGTGGCCGACCC 607
Db 542 GGGCAGGGCGGCTAGGGCGGCGCAGTCCGCCCTTGGCCCGCCAGCCAGTGGCCGACCC 601
QY 608 ACTCCCTGCCACCGCTCACAAGTATTATTGTTCCCAATGCTGGAGGGGCGCTTC 667
Db 602 ACTCCCTGCCACCGCTCACAAGTATTATTGTTCCCAATGCTGGAGGGGCGCTTC 661
QY 668 CAGATTGGGGCGCTGGGGTCCCGACATCCCTGTCCTCCATCCCGCTGGGCTGGCAGCC 727
Db 662 CAGATTGGGGCGCTGGGGTCCCGACATCCCTGTCCTCCATCCCGCTGGGCTGGCAGCC 721
QY 728 AGATTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGAGGCTCCAGACCCAGGGCAGT 787
Db 722 AGATTCTCCCTTAAGGAATTGACTTCAGCAGGGGTGGAGGCTCCAGACCCAGGGCAGT 785
QY 788 GTGGTGGAGGGGTGTTCCAAAGAGAGAGGCTGGTTCAGCAGA 829
Db 781 GTGGTGGAGGGGTGTTCCAAAGAGAGAGGCTGGTTCAGCAGAAA 821

RESULT 14
LOCUS BF753931
DEFINITION 603022938F1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5193728 5',
mRNA sequence.
ACCESSION BF753931
VERSION BF753931.1 GI:15745509
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 964)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/.
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM11484 row: k column: 09
High quality sequence stop: 876.
Location/Qualifiers
1..964
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5193728"
/lab_host="DH10B"

```


/clone lib="NIH_MGC_114"
 /note="Organ: Brain; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: EcoRV (destroyed); RNA source anonymous pool of 6
 male brains, age range 23-27 yo. Library is oligo-dT
 primed and directionally cloned (EcoRV site is destroyed
 upon cloning). Average insert size 1.5 kb, insert size
 range 1-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 019. Note:
 this is a NIH_MGC Library."

ORIGIN

Query Match 78.0%; Score 775.6; DB 12; Length 964;
 Best Local Similarity 93.5%; Pred. No. 2.1e-146;
 Matches 899; Conservative 0; Mismatches 49; Indels 14; Gaps 8;

QY 26 TGGCGGACGAGGAGAGTCCGCGCGCTGGGAGAGCGCATGAGCCGAGCTCAGGCC 85
 Db 1 TGGCGGACGAGGAGAGTCCGCGCGCTGGGAGAGCGCATGAGCCGAGCTCAGGCC 60

QY 86 GAGTGTACTTCAACCACTACATACGCCAGCTGGGAGCGGCCCGAGCGCAACA 145
 Db 61 GAGTGTACTTCAACCACTACATACGCCAGCTGGGAGCGGCCCGAGCGCAACA 120

QY 146 GCAGCAGTGGTGCAGGAGCGGAGCGCTGCCAGGCTCCGCTGCTGCACCTGC 205
 Db 121 GCAGCAGTGGTGCAGGAGCGGAGCGCTGCCAGGCTCCGCTGCTGCACCTGC 180

QY 206 TGGTGAAGCAGACGAGTCAAGCGCGCTCTCTGGCGGAGGAGATCACCCGA 265
 Db 181 TGGTGAAGCAGACGAGTCAAGCGCGCTCTCTGGCGGAGGAGATCACCCGA 240

QY 266 CCAAGAGAGGAGGCTGGAGTGTATCAAGCTATCATCCAGAGATCAAGTCGGAGAGG 325
 Db 241 CCAAGAGAGGAGGCTGGAGTGTATCAAGCTATCATCCAGAGATCAAGTCGGAGAGG 300

QY 326 AGGCTTTGAGTCTGAGCTCAGAGTTCAGGAGTGCAGTCCAGGAGCGGAGGAG 385
 Db 301 AGGCTTTGAGTCTGAGCTCAGAGTTCAGGAGTGCAGTCCAGGAGCGGAGGAG 360

QY 386 ACCTGGGTGCTTCAGCAGAGTGCAGTCAAGCCATTTCAAGAGCGCTGTTTGGC 445
 Db 361 ACCTGGGTGCTTCAGCAGAGTGCAGTCAAGCCATTTCAAGAGCGCTGTTTGGC 420

QY 446 TGGGACGCGGAGAGTATGAGCGGCGCGTGTTCAGGATTCGGGATCCACATCATCTCC 505
 Db 421 TGGGACGCGGAGAGTATGAGCGGCGCGTGTTCAGGATTCGGGATCCACATCATCTCC 480

QY 506 GCATGTAGTGGGTGGGAGCGCCAGGCGCTGCGCTCGGCGGAGGAGCGGCGGCGG 564
 Db 481 GCATGTAGTGGGTGGGAGCGCCAGGCGCTGCGCTCGGCGGAGGAGCGGCGGCGG 540

QY 565 CGGCCAGCTCCCTTGCCTGCGGAGCGGAGTGGCGGAGCGGCGGAGCGGCGGCGG 624
 Db 541 CGGCCAGCTCCCTTGCCTGCGGAGCGGAGTGGCGGAGCGGCGGAGCGGCGGCGG 600

QY 625 CACAGTATTTATTTTCCACAAATGGCTGGGAGGCGGCGGCTTCAGATTTGGGCGCTGG 684
 Db 601 CACAGTATTTATTTTCCACAAATGGCTGGGAGGCGGCGGCTTCAGATTTGGGCG 659

QY 685 GGTCCCACTCCCTGCTCCATCCAGTGGGAGTGGGAGCGGAGTTCCTTCAAGGA 744
 Db 660 GGTCCCACTCCCTGCTCCATCCAGTGGGAGTGGGAGCGGAGTTCCTTCAAGGA 718

QY 745 ATTCACTTCAGAGAGGAGTGGAGGCTCCAGACCCAGGAGGAGTGGTGGAGGAGGTT 804
 Db 719 ATTCACTTCAGAGAGGAGTGGAGGCTCCAGACCCAGGAGGAGTGGTGGAGGAGGTT 777

QY 805 CCAAGAGAGAGGCTGG-----TCAGCAGAGCGGCGGCGGCTGTCGCCAGGAGTGGAGGC 860
 Db 778 CCAAGAGAGAGGCTGGTTCAGCAGAGCGGCGGCGGCTGTCGCCAGGAGTGGAGGC 837

QY 861 ---AGACTCGAGGCGCGAATTTTCTAGTTAGGCGCAGCTCTCTGTTAGTTCGCAAG 917

Db 838 AAGCTCTGAGGCGCGCAATTTCTAGTTAGGC--AGCTCTCTCTGTCAGGTCGCAAG 895
 QY 918 GTGAACACTCATCCGCGAGCGCATGCGCCCTCTGAGCAACTGTGCGAGACCCCTTTTACCCCC 977
 Db 896 GTGAACACTCATCCGCGAGCGCATGTG-CCTTTGAGCAACTGTGCGGCCCTTTTACCCCCATT 954
 QY 978 AA 979
 Db 955 AA 956

RESULT 15
 BE797127
 LOCUS BE797127
 DEFINITION mRNA sequence.
 BE797127
 BE797127.1 GI:10218325
 EST.
 Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 874)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: DCTD/DTF
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCW793 row: o column: 04
 High quality sequence stop: 734.
 Location/Qualifiers
 1..874
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3941595"
 /tissue_type="small cell carcinoma"
 /cell_line="MGC3"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_7"
 /note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

FEATURES
 source

Query Match 77.7%; Score 772.2; DB 10; Length 874;
 Best Local Similarity 96.5%; Pred. No. 1e-145;
 Matches 822; Conservative 0; Mismatches 24; Indels 6; Gaps 3;

QY 24 GATGCGGACGAGGAGAGTCCGCGCGCTGGGAGAGCGCATGAGCCGAGCTCAGG 83
 Db 1 GATGCGGACGAGGAGAGTCCGCGCGCTGGGAGAGCGCATGAGCCGAGCTCAGG 60

QY 84 CCGAGTGTACTTCTCAACCACTACATCAAGCCAGCGGAGCGCGCCAGCGCA 143
 Db 61 CCGAGTGTACTTCTCAACCACTACATCAAGCCAGCGGAGCGCGCCAGCGCA 120

QY 144 CAGCAGCAGTGGTGGCAAAACGCGGAGGAGCGCTGCCAGGTCGCTCGACCT 203
 Db 121 CAGCAGCAGTGGTGGCAAAACGCGGAGGAGCGCTGCCAGGTCGCTCGACCT 180

ORIGIN

QY	204	GTGGTGAAGCA	CAGCAGCAGT	CACGGCGGCCCT	GTCTCTGGCGGAGGAAAGAT	CACCGG	263
Db	181	GTTGGTAAAGCA	CAGCAGCAGT	CACGGCGGCCCT	GTCTCTGGCGGAGGAAAGAT	CACCGG	240
QY	264	GACCAAGGAGGAGCC	CTGGAGCTGATCAACGGCTACAT	TCGAGAGATCAAGT	CGGGAGA	323	
Db	241	GACCAAGGAGGAGCC	CTGGAGCTGATCAACGGCTACAT	TCGAGAGATCAAGT	CGGGAGA	300	
QY	324	GGAGGACTTTGAGTCT	CTCGGCTCTCAGTTTCAGCGACT	GCAGCTCAGCCAAAGCC	CAGGGG	383	
Db	301	GGAGGACTTTGAGTCT	CTCGGCTCTCAGTTTCAGCGACT	GCAGCTCAGCCAAAGCC	CAGGGG	360	
QY	384	AGACCTGGGTGCTTT	CAGCAGAGGTCAGATG	CAGAGCCATTTGAAGAGC	CTCGTTTGC	443	
Db	361	AGACCTGGGTGCTTT	CAGCAGAGGTCAGATG	CAGAGCCATTTGAAGAGC	CTCGTTTGC	420	
QY	444	GTCGGGAGCGGGGAGAT	GCAGCGGGCCCGTGTTCACGGAT	TCCCGCATCCCATCATCTCT	503		
Db	421	GTCGGGAGCGGGGAGAT	GCAGCGGGCCCGTGTTCACGGAT	TCCCGCATCCCATCATCTCT	480		
QY	504	CGCAGTGAAGTGAAGG	TGGGAGCCACAGCCCTG	GGCTCGGGCAGGGCAGGCGCT	TAGG	563	
Db	481	CGCAGTGAAGTGAAGG	TGGGAGCCACAGCCCTG	GGCTCGGGCAGGGCAGGCGCT	TAGG	540	
QY	564	CGGGCAGCTCCCTCT	TCGCGCCACGAGTGGCCG	GAACCCCACTCCCTGCCACCGTC	623		
Db	541	CGGGCAGCTCCCTCT	TCGCGCCACGAGTGGCCG	GAACCCCACTCCCTGCCACCGTC	600		
QY	624	ACACAGTATTTATTT	GTTCCTCCCAATGTGTGGAGGGGGCCCT	TCAGATTGGGGGCGCCCTG	683		
Db	601	ACACAGTATTTATTT	GTTCCTCCCAATGTGTGGAGGGGGCCCT	TCAGATTGGGGGCGCCCTG	660		
QY	684	GGGTCCCCACATCCCT	GTCTCATCCCACTGGGGGTG	CGACCGACAGATTCTCCCTTAAGG	743		
Db	661	GGGTCCCACTCCCT	GTCTCATCCCACTGGGGGTG	CGACCGACAGATTCTCCCTTAAGG	720		
QY	744	AATTGACTTTCAGCGGG	TGGAGGCTCCACAGCC	CCAGGGCAGTGTGTGGAGGGGTGT	803		
Db	721	CATTGACTTTCAGCGGG	TGGAGGCTCCAGAAC	CCAGGGCAGTGTGTGGAGGGGTGT	779		
QY	804	TCCAAAGAGAGGCTT	GGTTCAGCAGAGCGGCCCGTGT	CTCCCGCAGGTGTGGAGGAGG	863		
Db	780	TCCAAAGAGAGGCT	GGATCCAGAGAGGCTGGAC	-----GCAAGCGCCCGGCTCCCGCAGGTGTGGAGGCGAC	834		
QY	864	CTCAGAGGCGGA	875				
Db	835	TCCAGAGGCGCAA	846				

Search completed: June 28, 2004, 18:01:43
Job time : 3452.82 secs